Российский государственный гуманитарный университет
Russian State University for the Humanities

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

Journal of Language Relationship

International Scientific Periodical

№ 17/2

Moscow 2019

Российский государственный гуманитарный университет
Институт языкознания Российской Академии наук

Вопросы языкового родства

Международный научный журнал

№ 17/2

Москва 2019

Журнал основан К. В. Бабаевым

Вопросы языкового родства: Международный научный журнал / Рос. гос. гуманитар. ун-т; Рос. акад. наук. Ин-т языкознания; под ред. В. А. Дыбо. — М., 2019. — № 17/2. — x + 84 с.

Journal of Language Relationship: International Scientific Periodical / Russian State University for the Humanities; Russian Academy of Sciences. Institute of Linguistics; Ed. by V. A. Dybo. — Moscow, 2019. — No. 17/2. — x + 84 p.

ISSN 2219-3820

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

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

Подписано в печать 10.12.2018. Формат 60×90/8.
Усл. печ. л. 10,9. Уч.-изд. л. 11,4. Заказ № 511.
Тираж 1050 экз.

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

Table of Contents / Содержание

Articles / Статьи

Contributors

Andrey Nikulin — postgraduate student, University of Brasília; andre.n.guzman@gmail.com

Giorgio Orlandi — independent researcher, thegaygenius@gmail.com

George Starostin — senior researcher, Institute of Oriental and Classical Studies, Higher School of Economics / Russian State University for the Humanities; gstarst1@gmail.com

Сведения об авторах

Никулин, А. В. — аспирант, Университет Бразилиа, andre.n.guzman@gmail.com

Орланди, Джорджио — независимый исследователь, thegaygenius@gmail.com

Старостин, Г. С. — старший научный сотрудник Института классического Востока и античности ВШЭ, заведующий кафедрой истории и филологии Дальнего Востока ИВКА РГГУ, 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
Miusskaya Square, 6
E-mail: gstarst@rinet.ru

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

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

Будущим авторам

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

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

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

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

Andrey NIKULIN
University of Brasília; andre.n.guzman@gmail.com

A reconstruction of Proto-Jê phonology and lexicon[1]

In this work, I examine the sound correspondences between Proto-Cerrado (Nikulin 2017) and Proto-Southern Jê (Jolkesky 2010) and offer a phonological reconstruction of Proto-Jê, the proto-language of the most diverse subgrouping within the Macro-Jê language stock. I reconstruct 11 consonants and 19 vowels for this proto-language. I also claim that */CrVC/ was the maximal syllable structure in Proto-Jê with some further restrictions on its complex onsets (only */pr, mr, kr, ŋr/ were allowed). I reconstruct a shielding allophony pattern to Proto-Jê, according to which nasal onsets would have had post-oralized allophones before oral nuclei. The discussion on Proto-Jê phonology is followed by a sample of Proto-Jê lexicon.

Keywords: Jê languages, Macro-Jê languages, language reconstruction, comparative method.

1. Introduction

Since the pioneer work of Davis (1966), which remains the only published attempt at a reconstruction of Proto-Jê to the present day, numerous descriptive studies on individual Jê languages have been carried out, contributing to a significant improvement in the state-of-the-art of Jê linguistics. Recent comparative work on low-level branches of Jê, such as Southern Jê, Northern Jê, Central Jê, and Cerrado (Wiesemann 1978, Jolkesky 2010, Nikulin & Salanova forthc., Nikulin 2016, 2017), has enabled the use of intermediate reconstructions in comparative studies on higher levels. In this paper, I offer a revised reconstruction of Proto-Jê, taking into account the research on Jê produced over the last 50 years.

I follow the principles of bottom-up reconstruction, which means that my reconstruction of Proto-Jê is based on the reconstructions of its daughter proto-languages, Proto-Cerrado (Nikulin 2017) and Proto-Southern Jê (Wiesemann 1978, Jolkesky 2010); the former, in turn, is based on the reconstructions of Proto-Northern Jê (Nikulin 2016, cf. also Nikulin & Salanova forthc.) and Proto-Central Jê (Nikulin 2017). This allows to filter out recent phonetic, morphological, and semantic innovations, which is especially important in light of the conservativeness traditionally attributed to Northern Jê (cf. Ribeiro & Voort 2010: 560, Carvalho 2016: 64, Pache 2018); it is well known that over-reliance on the data of a given

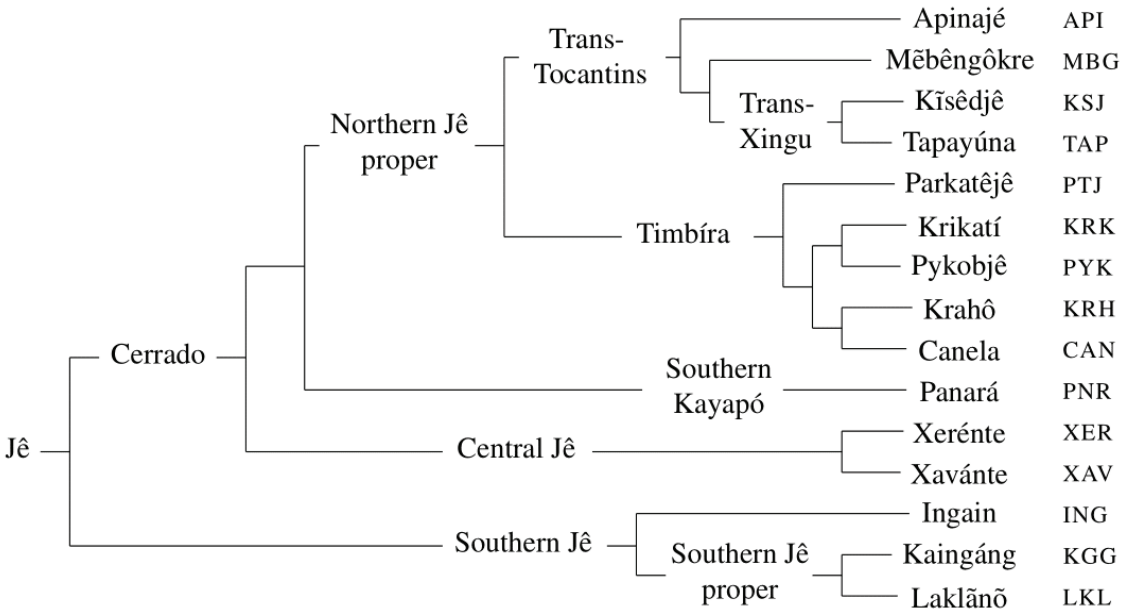
[1] I gratefully acknowledge the financial support of CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) during my doctoral studies. I also thank Andrés Pablo Salanova, Mário André Coelho da Silva, and Matthias Pache for valuable input that helped improve this paper. Furthermore, its text has benefited from the comments of Juho Pystynen, Mikhail A. Zhivlov, and Lev Michael, to whom I am grateful. I warmly thank my Jê-speaking friends, especially Bepo Metyktire and Eneida Xerente, for teaching me bits of their languages. Finally, I thank the audience at Amazónicas VI (Tabatinga), where a much earlier version of this paper was presented on May 27th, 2016.

I use the following abbreviations throughout the paper: PCerr = Proto-Cerrado, PCJ = Proto-Central Jê, PJ = Proto-Jê, PNJ = Proto-Northern Jê, PSJ = Proto-Southern Jê, ABL = ablative, ACT = active, ADESS = adessive, ADVT = advertence prohibitive, ALL = allative, CAUS = causative, COP = copula, DU = dual, ERG = ergative, HABIT = habitual, INSTR = instrumental, INTR = intransitive, LOC = locative, MALEF = malefactive, NEG = negative, NF = non-finite, NMLZ.AG = agent nominalization, PL = plural, SG = singular, STAT = stative, ATR = advanced tongue root.

language or language group has repeatedly led historical linguists to wrong reconstructive solutions.[2] It is important to emphasize that bottom-up reconstruction does not exclude occasional use of external material, which may be invoked in situations when internal data offer conflicting or incomplete evidence.

My working model of the Jê phylogenetic tree (without Xakriabá and Akroá, whose position within Central Jê is uncertain) is presented in Figure 1.[3]

Figure 1. Internal structure of the Jê family



The remainder of this paper is organized as follows. Section **2** contains an outline of the phonological systems reconstructible to Proto-Cerrado (subsection **2.1**) and Proto-Southern Jê (subsection **2.2**). In section **3**, I examine the sound correspondences between Proto-Cerrado and Proto-Southern Jê and come up with a reconstructive proposal for Proto-Jê. Onsets (subsection **3.1**), nuclei (subsection **3.2**), and codas (subsection **3.3**) are dealt with separately. A list of the most reliable Jê etymologies is provided in section **4**. Section **5** summarizes the paper and lists several unresolved issues in comparative Jê linguistics that should be subject to future research.

[2] One noteworthy example is the bias towards Finnic data in Uralic comparative studies, which prevented the scientific community from providing a correct account of several important vowel correspondences for decades; only very recently (Aikio 2012, 2015) did it become clear that Finnic vocalism is innovative in a number of respects. Another well known case was the over-reliance on Sanskrit data in mid-nineteenth century Indo-European comparative studies, which led scholars to a long-lasting misconception that Proto-Indo-European vocalism was identical to that of Sanskrit (Benware 1974: 66–81).

[3] The tree in Figure 1 was initially obtained through the application of StarlingNJ, a distance-based lexicostatistical method (see Kassian 2015: 5 for details), to 110-item wordlists of 16 extant Jê varieties (including several dialects of Kaingáng and Mẽbêngôkre but excluding Krikatí and Ingain). The annotated wordlists are scheduled to be published online at <http://starling.rinet.ru/new100> in 2019 as a part of the Global Lexicostatistical Database. For each node identified by the StarlingNJ algorithm, shared lexical, phonological and/or morphosyntactic innovations can be presented. For reasons of space, I do not discuss the internal classification of Jê in what follows, hoping to revisit the subject in future publications.

2. Proto-Cerrado and Proto-Southern Jê

In this section, I describe in detail the Proto-Cerrado (PCerr) and Proto-Southern Jê (PSJ) reconstructions adopted in this paper.

2.1. Proto-Cerrado

The Cerrado branch of the Jê family is constituted by two subbranches, Northern Jê[4] and Central Jê. I use the reconstruction of Proto-Cerrado by Nikulin (2017), according to which the phonological inventory of this proto-language comprised 12 consonants and 15 vowels. PCerr consonantal phonemes are shown in Table 1 below.

Table 1. Proto-Cerrado consonants.[5]

	labial	coronal	palatal	velar
oral stops	*p	*t	*c	*k
nasal stops	*m	*n	*ñ	*ŋ
approximants	*w	*r	*j	*ɰ[6]

In oral environments, underlying nasal onsets were realized as postoralized (**mb-*, **nd-*, **nĵ-*, **ŋg-*). In nasal environments, the contrasts **j*/**ñ* and **ɰ*/**ŋ* were neutralized in favor of **ñ* and **ŋ*, respectively. All these allophonic processes are reflected in the transcription system used in this work. The pronunciation of PCerr consonants in coda position is uncertain, but at least **m* is likely to have had a preoralized allophone *[-ᵇm] in oral environments (unless followed by an echo vowel, for which see below).

The 15 vowels of PCerr are listed in Table 2 below.

---

[4] For convenience, I will use "Northern Jê" and "Proto-Northern Jê" ("PNJ") as synonyms of "Northern Jê proper" and "Proto-Northern Jê proper" (that is, excluding Panará/Southern Kayapó) throughout this paper. I will refer to Northern Jê *lato sensu* as "Northern Jê–Panará" whenever necessary.

[5] In this work, I diverge from my earlier practice of representing linguistic data using the Unified Transcription System (UTS) and adhere to a modified version of the International Phonetic Alphabet (IPA) instead. The main differences between the transcription system used in this work and IPA are: *r* = IPA *ɾ*; *ñ* = IPA *ɲ*; *c* stands for any voiceless palatal, alveo-palatal or postalveolar stop, or affricate; *ĵ* stands for any voiced palatal, alveo-palatal or postalveolar stop or affricate; *x* stands for a voiceless alveo-palatal or postalveolar fricative; *ə, o, e, ə̃, õ, ẽ* are open-mid or simply mid; *ə̂, ô, ê* are close-mid; *ə, ə̂, ʏ, y* are central or back unrounded vowels (open-mid, close-mid, near-close, and close, respectively); *ĭ, ŭ, ŷ* are falling lowering diphthongs. In Karajá data, the circumflex stands for the [+ATR] feature. The apostrophe in PCerr reconstructed forms is used to indicate that the echo vowel is absent (see Nikulin 2017: 168–169 on this phenomenon).

[6] Although the reflexes of this segment in all Cerrado languages are stops (MBG *g*, API, TIM, TAP, KSJ, PNR, XER *k*, XAV ʔ), it patterns phonologically with PCerr **j* and **w* in a number of respects (see Nikulin & Salanova forthc. for discussion), suggesting that at least phonologically it belonged to the natural class of approximants in Proto-Cerrado. Fortition of approximants, especially in stressed syllables, is a recurrent trend across Cerrado languages: the development **w* > **b* is known from Proto-Northern Jê–Panará, and the development **j* > **ĵ* is hypothesized for Proto-Cerrado (in all environments), for Proto-Northern Jê–Panará (in stressed syllables), and for Proto-Trans-Tocantins (in unstressed syllables preceding **ə, *u*). That way, it does not seem entirely impossible that the segment in question was an approximant not only phonologically, but also phonetically in Proto-Cerrado.

Table 2. Proto-Cerrado vowels.

	oral			nasal		
	front unrounded	non-front unrounded[7]	back rounded	front unrounded	non-front unrounded	back rounded
close	*i	*y	*u	*ĩ	*ỹ	
close-mid	*ê	*ə̂	*ô			
open-mid	*e	*ə	*o	*ẽ	*ə̃	*õ
open		*a				

An already rich vocalic system reconstructible to Proto-Cerrado became even richer in Proto-Northern Jê, where **ã* and **ũ* emerged due to positional nasalization.

Apart from the monophthongs, at least two complex nuclei, **wa* and **ja*,[8] are reconstructible to Proto-Cerrado. Note that new important etymologies have been identified since the publication of the previous proposal (Nikulin 2017: 163), corroborating the reconstruction of the diphthongs in question. I list these etymologies in (1) below.

(1) New Cerrado etymologies involving a diphthong
   a. PCerr **cwa/*cwa-r*' 'to ask' > PNJ **cû/*cwə̂-r*, PCJ **wa/*wa-ri*
   b. PCerr **jwañ*' 'NMLZ.AG' > PNJ **ĵwə̂ñ*, PCJ **-kwaj // *-kwa*
   c. PCerr **jajwa/*jajwa-r*' 'to lay.PL, to spill' > PNJ **jaĵû/*jaĵwə̂-r*, PCJ **c-a(j)kwa/*c-a(j)kwa-ri*
   d. PCerr **ŋgrwa* 'moriche' > PNJ **ŋgrwa ~ *ŋgrû*, PCJ **wa:-bu* 'moriche stem' (< PCerr **ŋgrwa-pu* > PNJ **ŋgrwa-pu*)
   e. PCerr **ñõkwa* 'sternum' > PNJ **ñõkwa*, PCJ **ñõwa* 'in front of'
   f. PCerr **ñõrkwa* 'home' > PNJ **ñũrkwa*, PCJ **ñõrõwa*
   g. PCerr **krwat* 'beak' > PNJ **krwə̂t*, PCJ **wata // *wada*
   h. PCerr **rwaj-ci* 'rib' > PNJ **rwə̂-ci*, PCJ **waj-hi*
   i. PCerr **rwaj*' 'moon' > PNJ **mbyt-rwə̂*, PCJ **waj // *wa*
   j. PCerr **twam*' 'fat' > PNJ **twə̂m*, PCJ **wam // *wa*
   k. PCerr **a-ŋgja/*ŋgja-c* 'to enter.PL' > PNJ **a-ŋgî ~ *a-ŋgjê/*ŋgjê-c*, PCJ **ə̃-ĵa/*ĵa-ci* 'to enter.DU'[9]
   l. PCerr **ŋgja/*ŋgja-ñ*' 'to enter.CAUS.PL' > PNJ **ŋgî ~ *ŋgjê/*ŋgjê-ñ*, PCJ **ĵaj // *ĵa* 'to enter.CAUS.DU'
   m. PCerr **krja/*krja-r*' 'to bring up' > PNJ **krî/*kjê-r*, PCJ **ĵa/*ĵa-ri*

[7] It is uncertain whether non-front unrounded vowels were phonetically central or back.

[8] Nikulin (2017) adopted the reconstruction **jê*. The change in reconstruction is suggested by the near-symmetry of the reflexes of PCerr **ja* and **wa* in daughter languages. In Proto-Northern Jê, these two diphthongs yielded **jê* and **wə̂* in closed syllables (as well as in open syllables that go back to PCerr syllables ending in **-j*), but **jê ~ *î* and **wa ~ *û* in originally open syllables (with a so far unclear variation that is sometimes preserved in Proto-Northern Jê, see Nikulin 2017: 178–179). In Proto-Central Jê, the reflexes of these diphthongs are **ĵa* and **wa*, and any PCerr onset is eliminated before the diphthong (except that PCerr **jwa* > PCJ **kwa*).

[9] The etymologies (1k) and (1l) had been erroneously conflated in Nikulin 2017. Although related, they are two clearly distinct verbs. The same semantic and morphological relation holds for their singular counterparts: PCerr **a-jə/*jəp-r* 'to enter.SG' > PNJ **a-ĵə/*ĵə-r*, PCJ **ə̃-ĵe/*ĵeb-re*; PCerr **jə̲/*jə̲-r* 'to enter.CAUS.SG' > PNJ **ĵə/*ĵə-r*, PCJ **ñẽ/*ñẽ-rẽ*.

n. (?) PCerr **kwa/kwa-r'* 'to crack nuts' > PNJ **kukwə̂/*kukwə̂-r*, PCJ **wa/*wa-ri*
o. (?) PCerr **krwaj* 'parrot' > PNJ **krwə̂j*, PCJ **waj // *wa, *waj-hərə*

In Proto-Cerrado, the onset position could be filled with any single consonant or by a cluster formed by a peripheral (labial or velar) stop and **-r-*: **pr, *m(b)r, *kr, *ŋ(g)r*. Strikingly, the occurrence of onsetless syllables appears to have been very limited: only pretonic syllables that contained **a* or **i* as its nucleus could lack a consonantal onset.

Syllables could also have a coda in Proto-Cerrado, which could be followed or not by a so called *echo vowel* (only word-finally). In my notation, the occurrence of an echo vowel after a consonantal coda is unmarked, whereas the absence of a word-final echo vowel is marked with the apostrophe. The following codas can be reconstructed for Proto-Cerrado: **p, *t, *c, *k, *m, *n, *r, *j, *pr, *m', *n', *ñ', *r', *j', *pr'*. The complex codas **pr* and **pr'* occur exclusively in non-finite verb forms, when the non-finite suffix **-r* is adjoined to a verbal stem that ends in an underlying labial consonant. Very few examples are available that could substantiate the reconstruction of Proto-Cerrado **c, *n, *ŋ, *j, *n', *ñ'* in coda position.[10] The reconstruction of Proto-Cerrado **d*, as proposed by Nikulin (2017: 169), is even more problematic in that no **d* can be reconstructed for the onset position.

One major issue in Cerrado historical phonology remains entirely unresolved. Namely, some Central Jê stems unexpectedly contain nasal nuclei in stems whose cognates in Northern Jê (and in other Macro-Jê languages, whenever available) display oral nuclei. Examples include PCJ **ñẽ/*ñẽ-rẽ* 'to enter.CAUS.SG' (cf. PNJ **ĵə/*ĵə-r*), **c-ə̃/*c-ə̃-rĩ* 'to stand.CAUS.SG' (cf. PNJ **ĵa/*ĵa-r*), **ñə̃/*ñə̃m-rĩ* 'to weave' (cf. PNJ **cy/*cy-r*), **ĵatõ/*ĵatõ-rĩ* 'to send' (cf. PNJ **jando/ *jando-r*), **c-acõ/*c-acõm-rĩ* 'to hang' (cf. PNJ **janĵô/*janĵô-r*), **rẽ/*rẽ-mẽ* 'to leave.SG' (cf. PNJ **re*), **tõ/*tõ-rĩ* 'to urinate' (cf. PNJ **ij-tu/*tu-r*), **pə̃/*pə̃-rĩ* 'to kill.DU' (cf. PNJ **pa/*pa-r* 'to kill.SG'), **tə̃j // *tə̃* 'rain' (cf. PNJ **nda*). The origins of nasalization in Central Jê have not yet been identified, and it is uncertain how the respective PCerr forms should be reconstructed. I leave this question open and provisionally represent these unclear cases by underlining the nuclei in my PCerr reconstructions (e.g. **jə̲/*jə̲-r, *ja̲/*ja̲-r', *cy̲/*cy̲p-r', *jando̲/*jando̲-r', *janĵô̲/*janĵô̲p-r', *re̲, *tu̲/*tu̲-r', *pa̲/*pa̲-r', *nda̲j'*).

2.2. Proto-Southern Jê[11]

The reconstruction of PSJ consonantism does not pose serious difficulties, and the systems reconstructed by Wiesemann (1978) and Jolkesky (2010) are almost identical. I accept their reconstruction with minor changes in notation.

Although underlyingly nasal, voiced stops were subject to postoralization in oral environments (**mb-, *nd-, *nĵ-, *ŋg-*); this allophony pattern is active in both Southern Jê languages and is reconstructible to Proto-Cerrado. The preoralization of underlying nasal codas is also attested both in Kaingáng and Laklãnõ, though this process is not equally robust in all Southern Jê varieties: depending on the dialect and on the place of articulation, either the oral phase or the nasal phase may be omitted. Although clearly reconstructible to Proto-Southern Jê, I chose not to represent this allophony in my transcription.

[10] Known examples include PCerr **ŋgjac* 'to enter.PL.NF', **mbyn* 'tail', **poŋ* 'arrow, bamboo', **mbôŋ* 'lake', **krwaj* 'parrot', **mbjan'* 'husband', **ŋgja-ñ'* 'to enter.CAUS.PL.NF', **jwañ'* 'NMLZ.AG', **jəñ'* 'sweet', **ŋgwañ'* 'feather', maybe also **mbec* 'good' and **mbên* 'liquid'.

[11] In this paper, I do not consider Ingain data due to their ill-transcribed nature. Hereinafter the terms "Southern Jê" and "Proto-Southern Jê" are used in their narrow definition (that is, "Southern Jê proper" and "Proto-Southern Jê proper", excluding Ingain). Ingain data will need to be subjected to the procedure of *restitution* (Constenla Umaña 2000) before they can be meaningfully used in the reconstruction of Proto-Jê.

Table 3. Proto-Southern Jê consonants.

	labial	coronal	palatal	velar	glottal
oral stops	*p	*t	*c	*k	*ʔ
fricative		*θ			
nasal stops	*m	*n	*ñ	*ŋ	
approximants	*w	*r	*j		*h

Regarding the vocalic system of Proto-Southern Jê, I have previously argued (Nikulin 2015) that Jolkesky's (2010) reconstruction of PSJ vowels is seriously flawed and put forward a proposal backed up with morphophonological evidence. Note that while my reconstruction of PSJ oral vowels coincides with that of Wiesemann (1978), our reconstructions of PSJ nasal vowels are radically different.[12]

Table 4. Proto-Southern Jê vowels.

	oral			nasal		
	front unrounded	non-front unrounded	back rounded	front unrounded	non-front unrounded	back rounded
close	*i	*y	*u	*ĩ		*ũ
close-mid	*ê	*ə̂	*ô			
open-mid	*e	*ə	*o		*ə̃	
open		*a			*ã	

The phonotactic restrictions are identical in Kaingáng and Laklãnõ and are easily traced back to Proto-Southern Jê. The maximal syllable structure reconstructible to PSJ is **CrVC*. Just like in Proto-Cerrado, only labial and velar stops could form complex onsets with the rhotic: **pr*, **m(b)r*, **kr*, **ŋ(g)r*. As for the codas, only voiced segments could occur in that position; approximant codas were obligatorily followed by a non-phonemic echo vowel (phonetically a copy of the nucleus), whereas nasal codas did not trigger the occurrence of any vocalic segment. I leave PSJ echo vowels unmarked in my transcription because their occurrence and quality are fully predictable.

[12] Jolkesky (2010: 173) reconstructs yet another nasal vowel for Proto-Southern Jê, tracing the correspondence KGG *ə̃* ~ LKL *ũ* back to PSJ **ỹ* (**uĩ* in his transcription). Nikulin (2015: 286) reinterprets it as **õ* (**ɔ̃* in the transcription of the original paper). Note, however, that this sound correspondence occurs exclusively in grammatical morphemes, such as the interrogative pronoun KGG *nə̃* ~ LKL *nũ*, the situational marker KGG *kə̃* ~ LKL *kũ* and the subject marker KGG *wə̃* ~ LKL *vũ*. It now seems possible that this sound correspondence may have arisen as a conditioned development of PSJ **ɔ̃* (in this case Laklãnõ would have innovated by raising its usual reflex **õ* > *ũ*) or PSJ **ũ* (in this case Kaingáng would have innovated by unrounding and lowering its usual reflex **ũ* > *ə̃*) in unstressed grammatical morphemes.

3. Proto-Jê phonology

In this section, I examine the correspondences between Proto-Cerrado and Proto-Southern Jê onsets (**3.1**), nuclei (**3.2**), and codas (**3.3**), offering in each case a Proto-Jê reconstruction. Some Proto-Jê elements appear to have been preserved in only one constituent branch; in such cases, external evidence from other Macro-Jê languages has also been discussed.

3.1. Onsets

The correspondences between the onsets reconstructible for PCerr and PSJ are mostly trivial except for the palatal place of articulation, as shown in Table 5.

Table 5. Proto-Jê onsets and their reflexes in Proto-Cerrado and Proto-Southern Jê.

Proto-Jê		Proto-Cerrado		Proto-Southern Jê	
*/p/	**p*	*/p/	**p*	*/p/	**p*
*/m/	**mb, *m*	*/m/	**mb, *m*	*/m/	**mb, *m*
*/w/	**w*	*/w/	**w*	*/w/	**w*
*/pr/	**pr*	*/pr/	**pr*	*/pr/	**pr*
*/mr/	**mbr, *mr*	*/mr/	**mbr, *mr*	*/mr/	**mbr, *mr*
*/t/	**t*	*/t/	**t*	*/t/	**t*
*/n/	**nd, *n*	*/n/	**nd, *n*	*/n/	**nd, *n*
*/r/	**r*	*/r/	**r*	*/r/	**r*
*/c/	**c*	*/c/	**c*	*/θ/	**θ*
*/ñ/	**nĵ*	*/ñ/	**nĵ*	*/c/	**c*
*/j/	**j*	*/j/	**j*	*/j/	**j*
	**ñ*		**ñ*	*/n/	**n*
*/k/	**k*	*/k/	**k*	*/k/	**k*
*/ŋ/	**ŋg, *ŋ*	*/ŋ/	**ŋg, *ŋ*	*/ŋ/	**ŋg, *ŋ*
*/kr/	**kr*	*/kr/	**kr*	*/kr/	**kr*
*/ŋr/	**ŋgr, *ŋr*	*/ŋr/	**ŋgr, *ŋr*	*/ŋr/	**ŋgr, *ŋr*
*/ø/	**ø* (?)	*/ɯ̨/	**ɯ̨, *ŋ*	?	?

Most of the correspondences in Table 5 are identity correspondences, thus posing no difficulty for the reconstruction of the respective Proto-Jê segments. For reasons of space, the identity correspondences between Proto-Cerrado and Proto-Southern Jê onsets will not be examined in this section; numerous examples are available in Section 4.

Regarding the origin of the infrequent PCerr **ɰ*, it is unknown which segment should be reconstructed to Proto-Jê due to the fact that the tokens that contain this phoneme in Proto-Cerrado lack known cognates in Southern Jê. Nevertheless, it is certain that these tokens are of Jê origin, given that they have secure external cognates in other Macro-Jê languages, where PCerr **ɰ* corresponds to glottal segments (Maxakalí *h*; Krenák *h*; Jabutí *ʔ*; cf. Nikulin & Silva, forthc.). For lack of a better solution, I reconstruct **∅* (> PCerr **ɰ*) for Proto-Jê, but it should be kept in mind that other possibilities (e.g. **ɰ* or **h*) cannot be totally ruled out at present. Note that not a single onsetless monosyllabic root morpheme reconstructs to Proto-Cerrado, making it possible to stipulate an epenthetical origin for PCerr **ɰ*.

Palatal onsets. I claim that Proto-Cerrado is conservative with respect to the palatal onsets of Proto-Jê. Conversely, Proto-Southern Jê would have undergone a chain shift: as shown in (2–3) below, PJ **nĵ* > PSJ **c*, whereas PJ **c* > **θ*.

(2) Proto-Jê **nĵ* > Proto-Southern Jê **c*
   a. PJ **nĵy* ‘bee, bumblebee’ > PCerr **nĵy*, PSJ **cy*
   b. PJ **nĵô* ‘to hang.SG’ > PCerr **nĵô*/**nĵôp-r’*, PSJ **cə*/**ca-ŋ* (vb. **cə-m*)
   c. PJ **nĵu* ‘to heat, to dry’ > PNJ **nĵu*/**nĵu-ñ* ‘to dry in the sun’; PSJ **cu-r* (vb. **cu-n*) ‘to heat by the fire’, **jaŋ-cu*/**jaŋ-cu-ŋ, *jaŋ-cu-r* (vb. **jaŋ-cu-n*) ‘to roast’
   d. PJ **nĵê* ‘to tickle’ > PNJ **nĵê*/**nĵê-r*, PSJ **ku-cê*/**ku-ce-ŋ*
   e. (?) PJ **nĵa* ‘to bite’ > PCerr **nĵa*/**nĵa-r’*, PSJ **cã* ‘game *(animals)*; war’
   f. (?) PJ **nĵʊm’* ‘dirty’ > PNJ **nĵũm* ‘dirty’, PSJ **coŋ* ‘egg yolk; larvae’s food’
   g. (?) PJ **nĵi* ‘mother’ > PNJ **nĵi*, PSJ **ci* ‘old’

(3) Proto-Jê **c* > Proto-Southern Jê **θ*
   a. PJ **cy* ‘to weave’ > PCerr **cy*/**cyp-r’*, PSJ **θy*
   b. PJ **cym* ‘seed’ > PCerr **cym*, PSJ **θy* (vb. **θy-n*)
   c. PJ **côj* ‘leaf’ > PCerr **côj’*, PSJ **θej* (vb. **θe-ñ*)
   d. PJ **jacê ~ *jVmcê* ‘nest’ > PCerr **jacê*, PSJ **jəŋθê* (vb. **jəŋθe-ñ*)
   e. PJ **-cĩ(C)* ‘bandage, sling’ > PNJ **ja-cĩ, *ka-cĩ*, PSJ **-θĩ* (vb. *-θĩ-n*)

Note that PJ **c-* is also reconstructible as the allomorph of the third person marker used with **j*-initial stems. As suggested by Ribeiro (2004: 95–96, 2011: 109, fn. 8), Proto-Southern Jê appears to have fossilized an earlier third person prefix **c-* in some stems.[13]

   f. PJ **jô*, 3 (= third person) **c-ô* ‘bitter’ > PCerr **jô*, PSJ **θə* (vb. **θa-ŋ*)
   h. PJ **jara*, 3 **c-ara* ‘wing; armpit’ > PNJ **jara* ‘wing’, **jara-kre* ‘armpit’, PSJ **θãr* ‘wing’ (but **jãrã* ‘armpit’, vb. **jãrã-n*)

[13] It is worth observing that the stems that underwent this fossilization process are those that typically have an inanimate internal argument. This semantic peculiarity may have contributed to an elevated frequency of occurrences of these stems with a third person prefix in Proto-Jê discourse when compared to other **j*-initial stems, thus enabling the reanalysis. Another comment to be made is that this fossilization trend continued in Kaingáng even after the dissolution of Proto-Southern Jê: cf. PJ **ji*, 3 **c-i* ‘to lay.SG’ > PSJ **ji*, 3 **θ-i* > KGG *ɸi*, LKL *ji*, 3 *ð-i*. In other instances, on the contrary, Kaingáng generalized the uninflected form: PSJ **jãmã-ŋ*, 3 **θ-ãmã-ŋ* ‘to listen’ > KGG *jẽmẽ-ŋ*, LKL *ñãmã-ŋ*, 3 *ð-ãmã-ŋ*; PSJ **jô*, 3 **θ-ô* ‘for’ > KGG *jô*, LKL *jô*, 3 *ð-ô* (Wiesemann 1978: 209).

i. PJ **jʊ*, 3 **c-ʊ* ‘pus’ > PCerr **jup-r*, PSJ **θô* (vb. **θo-m*)
j. PJ **jê*, 3 **c-ê* (~ **jên*, 3 *c-ên*) ‘thread, knot, cloth’ > PNJ **ĵê*, PSJ vb. **θe-n* ‘to spin thread’
k. PJ **ja*, 3 **c-a* ‘to put vertically.SG’ > PCerr **ja̲/*ja̲-r*’, PSJ **θã* (vb. **θã-ŋ*) ‘to be put vertically’

In addition, Proto-Southern Jê depalatalized PJ **ñ* (the nasal allophone of */j/) to PSJ **n*, merging it with PSJ **n* < PJ **n*, as shown in (4) below.

(4) Proto-Jê **ñ* > Proto-Southern Jê **n*
a. PJ **ñĩ* ‘meat’ > PCerr **ñĩ*, PSJ **nĩ*
b. PJ **ñỹ* ‘to sit.SG’ > PCerr **ñỹ/*ñỹp-r*, PSJ **nĩ* (vb. **nĩ-m*)
c. PJ **ñĩjə̂* ‘smoke’ > PCJ **ñĩ̂je*, PSJ **nĩjə* (vb. **nĩja-ŋ*)
d. PJ **ñĩja* ‘nose’ > PCerr **ñĩja-kre*, PSJ **nĩjã*
e. PJ **ñũmjê(C)* ‘female breast’ > PNJ **ñõmĵê*, PSJ **nũŋjê* ‘female breast, milk’
f. PJ **ũt/*ñũt* ‘to sleep’ > PCerr **ŋõt/*ñõt*, PSJ **nũr*
g. PJ **ñĩ(m)-* ‘hand (*in compounds*)’ > PCerr **ñĩ(m)-kra* ‘hand’, PSJ **nĩ-* (as in **nĩ-pê* ‘to wash hands’)

The fact that Proto-Cerrado retains the original configuration of the inventory at the palatal point of articulation is suggested by the typology of the directionality of sound changes (from a cross-linguistic perspective, it is more plausible to posit a fricativization development such as **c* > **θ* than an unconditional fortition of a fricative), as well as by external data from Maxakalí, Krenák, and Karajá, in which languages PJ **c* corresponds to stops or affricates (Maxakalí *tɕ*; Krenák *kj*; Karajá *ɗ*, palatalized *tʃ*; cf. Nikulin & Silva, forthc.).

Word-medial clusters. Sometimes voiced segments are found in Proto-Southern Jê stems where, judging by their Proto-Cerrado cognates, voiceless consonants would be expected. I hypothesize that these apparently aberrant correspondences instantiate a sound law according to which voiceless stops became voiced (*/t, c/ > */n, j/) Proto-Southern Jê if preceded by an etymological coda (the coda itself may disappear). This rule ceased to be productive and no longer applies in new formations (compounds, reduplicated plurals etc.). Some examples are provided in (5).

(5) Voiceless consonants yield nasals in clusters in Proto-Southern Jê
a. PJ **ñũctə* ‘tongue’ > PCerr **ñõjto*, PSJ **nũnã*
b. PJ **k(r)Vmtym*’ ‘capybara’ > PCerr **kumtym*’, PSJ **kryŋndyŋ*
c. PJ **kapnĵa* ‘to chew’ > PCerr **kapnĵa/*kapnĵa-r*’, PSJ **kajã*
d. PJ **nĵê* ‘to tickle’ > PNJ **nĵê/*nĵê-r*, PSJ **ku-cê/*ku-ce-ŋ*, pl. **ku-ŋ-jê/*ku-ŋ-je-ŋ*

In a number of instances, it is unknown whether the Southern Jê words have a Jê etymology, but alternations attested in derived forms appear to stem from the same diachronic process:

e. PSJ **ku-cir* (vb. **ku-cin*) ‘to be roasted in fire’ → pl. **ku-ŋ-jir* (vb. **ku-ŋ-ji-n*)
f. PSJ **ku-côr* (vb. **ku-con*) ‘to be pierced’ → pl. **ku-ŋ-jôr* (vb. **ku-ŋ-jo-n*)
g. PSJ **cin* ‘to glean’ → antipassive **jəŋ-jin*
h. PSJ **cê/*ce-ŋ* ‘to tie’ → **kaŋ-jê/*kaŋ-je-ŋ*, pl. **kyŋ-jê/*kyŋ-je-ŋ* ‘to tie, to make a knot’; **jyŋ-jê/*jyŋ-je-ŋ* ‘to mend clothes’; **ãŋ-jê/*NP *jəŋ-jê* ‘trap with a maize bait’, (?) **krĩŋ-jê/*krĩŋ-je-ŋ* ‘to scratch; to go berserk’

3.2. Nuclei

The correspondences between PCerr and PSJ nuclei are significantly less straightforward than those between the onsets of these reconstructed languages. They are listed in Table 6 below; tentative reconstructions of Proto-Jê segments are included.

Table 6. Proto-Jê nuclei and their reflexes in Proto-Cerrado and Proto-Southern Jê.[14]

PJ	PCerr	PSJ
*e	*e	*e
*ê	*ê	*ê
*ɪ	*ja	*ê
*i	*i	*i

PJ	PCerr	PSJ
*a	*a	*ã
*ə	*o	*ã
*ə̂	*ə̂	*ə
*ʏ	*ə	*ə̂
*y	*y	*y

PJ	PCerr	PSJ
*o	*wa	*a
*ô	*ô	*ə
*ʊ	*u	*ô
*u	*u	*u

PJ	PCerr	PSJ
*ẽ	*ẽ	*ĩ
*ĩ	*ĩ	*ĩ

PJ	PCerr	PSJ
*ə̃	*ə̃	*ə̃
*ỹ	*ỹ, *ə̃†	*ĩ

PJ	PCerr	PSJ
*õ	*õ	*ə̃
*ũ	*õ	*ũ

† **ỹ* after a palatal onset, **ə̃* elsewhere

As can be seen from Table 6, I reconstruct a system with five contrastive vowel heights for Proto-Jê, in contrast with the systems of Proto-Cerrado and Proto-Southern Jê with only four vowel heights. Although systems with five contrastive heights are cross-linguistically rare (Ladefoged & Maddieson 1990: 95–96),[15] it is of notice that one modern Macro-Jê language, Karajá, has also been documented to have five contrastive vowel heights (Ribeiro 2012: 86), though in featural terms Karajá vowel heights are better understood as an interaction between a ternary height contrast and a privative [ATR] feature.

If this reconstruction is accepted, the evolution of PCerr and PSJ nuclei can be described in the following way.

- **From Proto-Jê to Proto-Cerrado** (ordering uncertain):
  (a) chain shift: **ʏ* > **ə*; **ə* > **o*; **o* > **wa*;
  (b) diphthongization of **ɪ* > **ja*;
  (c) raising of **ʊ* > **u*;
  (d) lowering of **ỹ* > **ə̃* (except after palatals) and of **ũ* > **õ* (unconditionally).

[14] I remind the reader that the transcription system adopted here differs from IPA in important ways (see fn. 5). Most importantly, the characters <ʏ, y, ỹ> stand here for *unrounded* non-front (central or back) vowels.

[15] Known cases include Austro-Bavarian as spoken in Amstetten (Traunmüller 1982) and Kensiu (Bishop 1996).

- **From Proto-Jê to Proto-Southern Jê** (order *a* → *b* → *c*; ordering of *d* and *e* uncertain)**:**
  (a) lowering of **ə* > **a*;
  (b) unrounding of **ô*/**o*/**õ* > **ɔ̂*/**ə*/**ɔ̃*;
  (c) lowering of near-close vowels: **ɪ* > **ê*; **ʊ* > **ô*; **ʏ* > **ɔ̂* (accompanied by a push chain lowering: **ɔ̂* > **ə*; **ə* > **a*; **a* > **ã*);
  (d) raising of **ẽ* > **ĩ*;
  (e) fronting of **ỹ* > **ĩ*.

I do not reject other possible interpretations of the sound correspondences exposed above; the scenario detailed here is the most parsimonious I was able to find. Future studies of external correspondences will undoubtedly contribute to corroborate or reject at least some of my claims.

It must be noted that Proto-Southern Jê close-mid and open-mid oral vowels are further lowered (**ɔ̂, *ô, *ê, *ə, *o, *e* > **ə, *o, *e, *a, *a, *a*, respectively) in syllables with nasal codas as a result of a synchronically active process. For space reasons, this process cannot be discussed here at length; the reader is referred to Nikulin (2015) for details.

Cognate sets that instantiate the aforementioned sound correspondences are provided below in (6–24).

(6) Proto-Jê **a* > Proto-Cerrado **a*, Proto-Southern Jê **ã*
  a. PJ **par* 'foot' > PCerr **par*, PSJ **pãn*
  b. PJ **pañ* 'arm, branch' > PCerr **paj*', PSJ **pã*
  c. PJ **mba* 'to hear, to understand' > PCerr **mba*/**mba-r*', PSJ **mã*/**mã-ŋ*
  d. PJ **mba* 'liver' > PCerr **mba*, PSJ **tə̃-mã*
  e. PJ **-mbaŋ*' 'to be afraid' > PCerr **pᵊ-mba*, PSJ **ka-mãŋ*, **mũ-mãŋ*
  f. PJ **wa* 'to walk' > PNJ **wa* 'to walk, to live', PSJ **wã* 'STAT' (aspect marker)
  g. PJ **ra* 'to hit' > PNJ **ku-ra*/**ku-ra-ñ*, PSJ **rã*/**rã-ŋ*/**rã-n*
  h. PJ **kapnĵa* 'to chew' > PCerr **kapnĵa*/**kapnĵa-r*', PSJ **kajã*
  i. PJ **jara* 'armpit; wing' > PNJ **jara* 'wing', **jara-kre* 'armpit', PSJ **jãrã* (vb. **jãrã-n*) 'armpit', **θãr* 'wing'
  j. PJ **ja* 'to stand.SG' > PCerr **ja*/**ja-m*', PSJ **jã*/**jã-ŋ*
  k. PJ **ja* 'to put vertically.SG' > PCerr **ja̲*/**ja̲-r*', PSJ **θã* (vb. **θã-ŋ*) 'to be put vertically'
  l. PJ **ñĩja* 'nose' > PCerr **ñĩja-kre*, PSJ **nĩjã*
  m. PJ **jañ* 'to eat.INTR, food' > PCJ **caj* // **ca*, PSJ **jã*/**jã-ŋ*/**jã-n* (vb. **jã-n*)
  n. PJ **kaj*' 'basket' > PNJ **kac*, PSJ **kãj* (vb. **kãñ*)
  o. PJ **kra* 'offspring' > PCerr **kra*, PSJ **krã* (vb. **krã-ŋ*)
  p. PJ **jandə* 'to send' > PCerr **jando̲*/**jando̲-r*', PSJ **jãnã*/**jãnã-ŋ*
  q. PJ **jarê* (~ **jarên*) 'root' > PNJ **jarê*, PSJ **jãrê* (vb. **θ-are-n*)[16]

[16] In examples (6q–s), one could be tempted to reconstruct PSJ **-a-* rather than PSJ **-ã-* in unstressed (non-final) syllables, given that the vowel in question is reflected as Kaingáng *õ* (*jõrê* 'root', *jõn-kõ* 'door', *kõ-ka* 'wind') rather than *ẽ* (usually PSJ **a, *ã* yield Kaingáng *õ, ẽ,* respectively). However, Laklãnõ data (*jãrê* 'root', *ð-ãn-ka* 'door') show that PSJ **ã* must be reconstructed for these tokens. I propose to consider the following sound law: PSJ **ã* > KGG *ẽ* in stressed (final) syllables as well as in unstressed (non-final) syllables if the stressed syllable also contains PSJ **ã* > KGG *ẽ* (cf. PSJ **jãnã* > KGG *jẽnẽ* 'to send'; PSJ **jãrã* > KGG *jẽrẽ* 'armpit'). Conversely, PSJ **ã* > KGG *õ* in unstressed (non-final) syllables if the stressed syllable contains a different vowel (cf. PSJ **jãrê* > KGG *jõrê* 'root'; PSJ **jãn-ka* > KGG *jõn-kõ* 'door'; PSJ **jarə* > KGG *jõra* 'saliva'). Note that *õ (~ ã)* and *ẽ* merged in all dialects other than Paraná Kaingáng and South-Eastern Kaingáng (Wiesemann 1978: 203–204); in some varieties, a new contrast between *õ* and *ẽ* emerged, whereby the use of *õ* is associated with round, compact objects and the use of *ẽ* is associated with long, diffuse objects (D'Angelis 2002). The Kaingáng forms given in this paper are from the Paraná dialect, where the etymological contrast is retained.

r. PJ **jar-ko* ‘mouth’ > PCerr **jad-kwa*, PSJ **jãn-ka* ‘door’
s. PJ **kôk ~ *ka-kôk* ‘wind’ > PCerr **kôk, *ka-kôk*, PSJ **kã-kə* (vb. **kã-kə-n*)

(7) Proto-Jê **ə* > Proto-Cerrado **o*, Proto-Southern Jê **ã*
a. PJ **ñũctə* ‘tongue’ > PCerr **ñõjto*, PSJ **nũnã*
b. PJ **jandə* ‘to send’ > PCerr **jando/*jando-r’*, PSJ **jãnã/*jãnã-ŋ*
c. PJ **ndəm* ‘eye’ > PCerr **ndom*, PSJ **kə-nã* (vb. **kə-nã-n*) ‘eye, fruit’
d. PJ **mbrə* ‘ashes’ > PCerr **mbro*, PSJ **mrã*
e. PJ **tə* ‘to fly.SG’ > PCerr **to/*top-r*, PSJ **tã* (vb. **tã-m*)

(8) Proto-Jê **ə̂* > Proto-Cerrado **ə̂*, Proto-Southern Jê **ə*
a. PJ **mbə̂* ‘to grab’ > PCerr **mbə̂/*mbə̂-r’*; PSJ **mbə/*mba-ŋ, *mbə-r* (vb. **mba-n*) ‘to hold, to carry (a short object)’
b. PJ **tə̂t* ‘strong, hard’ > PCerr **tə̂t*, PSJ **tər* (vb. **tan*)
c. PJ **jə̂*, 3 **c-ə̂* ‘bitter’ > PCerr **jə̂*, PSJ **θə* (vb. **θa-ŋ*)
d. PJ **ñĩjə̂* ‘smoke’ > PCJ **ñĩ̂je*, PSJ **nĩjə* (vb. **nĩja-ŋ*)

(9) Proto-Jê **ʏ* > Proto-Cerrado **ə*, Proto-Southern Jê **ə̂*
a. PJ **prʏ* ‘coal, ember’ > PCerr **prə* ‘ember’, PSJ **prə̂-ñ*
b. PJ **jʏ* ‘urine’ > PCerr **jə* ‘urine’, PSJ **jə̂-ñ* ‘to urinate’

(10) Proto-Jê **y* > Proto-Cerrado **y*, Proto-Southern Jê **y*
a. PJ **pry (~ *pryn)* ‘road’ > PCerr **pry*, PSJ **ã-pry/*NP jə-pry*
b. PJ **mbyt* ‘celestial body’ > PCerr **mbyt* ‘sun’, Ingain *pyr* ‘moon’
c. PJ **mbyn* ‘tail’ > PCerr **mbyn*, PSJ **mby*
d. PJ **ty* ‘to die’ > PCerr **ty(r)/*tyk*, PSJ **ty* ‘to die.STAT’
e. PJ **tyk* ‘black, dark’ > PCerr **tyk* ‘black’, PSJ **ku-ty* (vb. **ku-ty-ŋ*) ‘dark, night’
f. PJ **k(r)Vmtym’* ‘capybara’ > PCerr **kumtym’*, PSJ **kryŋndyŋ*
g. PJ **cy* ‘to weave’ > PCerr **cy/*cyp-r’*, PSJ **θy*
h. PJ **cym* ‘seed’ > PCerr **cym*, PSJ **θy* (vb. **θy-n*)
i. PJ **nĵy* ‘bee, bumblebee’ > PCerr **nĵy*, PSJ **cy*

(11) Proto-Jê **o* > Proto-Cerrado **wa*, Proto-Southern Jê **a*
a. PJ **tom’* ‘fat’ > PCerr **twam’*, PSJ **taŋ*
b. PJ **roñ* ‘celestial body’ > PCerr **rwaj’* ‘moon’, PSJ **ra* ‘sun’
c. PJ **jo* ‘tooth’ > PCerr **jwa*, PSJ **ja*
d. PJ **kVjo* ‘salt’ > PCerr **kVjwa*, PSJ **kəja* (vb. **kəja-ŋ*) ‘salty, sour’
e. PJ **jar-ko* ‘mouth’ > PCerr **jad-kwa*, PSJ **jãn-ka* ‘door’
f. PJ **kVñko* ‘sky’ > PCerr **kəjkwa*, PSJ **kəñka*
g. PJ **ŋgoñ’* ‘feather, hair’ > PNJ **ŋgwə̂ñ* ‘feather’, PSJ **ŋgañ* ‘animal hair’

(12) Proto-Jê **ô* > Proto-Cerrado **ô*, Proto-Southern Jê **ə*
a. PJ **pôc* ‘to leave.PL’ > PCerr **pôc*, PSJ **pə* (vb. **pə-n*)
b. PJ **wô* ‘to untie’ > PNJ **bô/*bô-ñ*, PSJ **kə-ɯə/*kə-ɯa-ŋ/*kə-ɯa-n* ‘to let loose, to untie’
c. PJ **nĵô* ‘to hang.SG’ > PCerr **nĵô/*nĵôp-r’*, PSJ **cə/*ca-ŋ* (vb. **cə-m*)
d. PJ **jô* ‘to tear, to rip’ > PNJ **ka-ĵô/*ka-ĵô-ñ*; PSJ **jə/*ja-ŋ, *jə-r* (vb. **ja-n*)
e. PJ **kôp* ‘fly, mosquito’ > PCerr **kôp* ‘fly’, PSJ **kə* ‘mosquito’
f. PJ **kôm* ‘tree, horn’ > PCerr **kôm*, PSJ **kə* ‘tree’, **nĩ-kə* ‘horn’
g. PJ **kôk ~ *ka-kôk* ‘wind’ > PCerr **kôk, *ka-kôk*, PSJ **kã-kə* (vb. **kã-kə-n*)
h. PJ **ŋgô* ‘louse’ > PCerr **ŋgô*, PSJ **ŋgə* (vb. **ŋga-n*) ‘louse, flea’

(13) Proto-Jê **ʊ* > Proto-Cerrado **u*, Proto-Southern Jê **ô*
- a. PJ **mbrʊm* 'ant' > PNJ **mbrũm*, PSJ **mbroŋ-jy* 'Argentine ant' (with nasal lowering)
- b. PJ **tʊm* 'old' > PCerr **tum* 'old, tall'; PSJ **toŋ* 'dry (*of plants*)' (with nasal lowering)
- c. PJ **pondʊ* 'bad *or* crooked' > PNJ **pundu* 'bad', PSJ **pandô* (vb. **pando-n*) 'crooked'
- d. PJ **jʊ*, 3 **c-ʊ* 'pus' > PCerr **jup-r*, PSJ **θô* (vb. **θo-m*)
- e. PJ **jʊm* 'father' > PNJ **ĵũ(m)*, PSJ **joŋ* (with nasal lowering)
- f. PJ **kʊ* 'to eat' > PCerr **ku*/**ku-r*', PSJ **kô* 'to eat, to use'

(14) Proto-Jê **u* > Proto-Cerrado **u*, Proto-Southern Jê **u*
- a. PJ **tu* 'to carry' > PCerr **tu*/**tu-r*', PSJ **tu* 'to carry on one's back, to wear'
- b. PJ **ru* 'to transport liquid or to pour' > PNJ **ru*/**ru-ñ* 'to spill, to pour', PSJ **ru*/**ru-ŋ*/ **ru-n* 'to fetch water'
- c. PJ **peju* 'to hide' > PNJ **puĵu*/**puĵu-r*, PSJ **peju*
- d. PJ **nĵu* 'to heat, to dry' > PNJ **nĵu*/**nĵu-ñ* 'to dry in the sun'; PSJ **cu-r* (vb. **cu-n*) 'to heat by the fire', **jaŋ-cu*/**jaŋ-cu-ŋ, *jaŋ-cu-r* (vb. **jaŋ-cu-n*) 'to roast'

(15) Proto-Jê **e* > Proto-Cerrado **e*, Proto-Southern Jê **e*
- a. PJ **mbec* 'good' > PCerr **mbec*, PSJ **mbe* 'HABIT', **ã-mbe* (vb. **ã-mba-n*) 'good weather'
- b. PJ **mbre* > PNJ **mbre* 'sibling-in-law', PSJ **jə-mbre* 'cousin, son-in-law'
- c. PJ **re* 'to leave, to abandon' > PCerr **re̲*, PSJ **re*/**ra-ŋ*
- d. PJ **kre* 'hole' > PCerr **kre*, PSJ **kre*
- e. PJ **ŋgre* 'egg' > PCerr **ŋgre*, PSJ **ŋgre* 'egg, penis'
- f. PJ **ŋgre* 'to dance' > PCerr **ŋgre̲*/**ŋgre̲-r*, PSJ **ŋgre* (vb. **ŋgre-n*)

(16) Proto-Jê **ê* > Proto-Cerrado **ê*, Proto-Southern Jê **ê*
- a. PJ **mbên* 'liquid' > PCerr **mbên*, PSJ **mbê* (vb. **mbe-n*)
- b. PJ **jarê* (~ **jarên*) 'root' > PNJ **jarê*, PSJ **jarê* (vb. **θare-n*)
- c. PJ **-rê(C)* 'to weed' > PNJ **ka-rê*/**ka-rê-r*, PSJ **ku-rê*/**ku-re-ŋ* 'to weed, to cut clean'
- d. PJ **jacê ~ *jVmcê* 'nest' > PCerr **jacê*, PSJ **jəŋθê* (vb. **jəŋθe-ñ*)
- e. PJ **nĵê* 'to tickle' > PNJ **nĵê*/**nĵê-r*', PSJ **ku-cê*/**ku-ce-ŋ*
- f. PJ **jê* (~ **jên*) 'thread, knot, cloth' > PNJ **ĵê*, PSJ vb. **θe-n* 'to spin thread'
- g. PJ **kêt* 'NEG' > PNJ **kêt*, PSJ **kêr* 'ADVT'

(17) Proto-Jê **ɪ* > Proto-Cerrado **ja*, Proto-Southern Jê **ê*
- a. PJ **mbɪn* 'husband' > PNJ **mbjên*, PSJ **mben* (with nasal lowering)
- b. PJ **kɪj* 'pit' > PNJ **kî* 'earth oven', PSJ **kêj* 'tomb'
- c. PJ **kɪ ~ *kɪj ~ *kɪj* 'to split' > PNJ **ko-kî*/**ko-kjê-r*, PSJ vb. **ke-ñ*
- d. PJ **krɪñ* 'thigh' > PCerr **k(r)jaj*', PSJ **krê*
- e. PJ **ŋgɪ* 'to enter.PL' > PCerr **a-ŋgja*/**ŋgja-c*, PSJ **ŋgê* (vb. **ŋgê-m*)

(18) Proto-Jê **i* > Proto-Cerrado **i*, Proto-Southern Jê **i*
- a. PJ **peti* 'to dream' > PNJ **pyti*/**pyti-r*, PSJ **peti*/**peti-ŋ*
- b. PJ **ji* 'to lay.SG' > PCerr **ji*/**ji-r*', PSJ **ji*/**ji-ŋ*
- c. PJ **jiji ~ *ñĩji* 'name' > PCerr **ñĩji*, PSJ **jiji ~ *jyjy* (vb. **jiji-n ~ *jyjy-n*)

(19) Proto-Jê **ə̃* > Proto-Cerrado **ə̃*, Proto-Southern Jê **ə̃*
- a. PJ **prə̃m* 'hungry, to want' > PCerr **prə̃m*', PSJ **prə̃ŋ* 'hungry, year'
- b. PJ **tə̃* 'INSTR' > PCerr **tə̃*, PSJ **tə̃* 'ERG, INSTR'
- c. PJ **nə̃* 'mother' > PCerr **nə̃*, PSJ **nə̃*
- d. PJ **ŋə̃(C)* 'to push against, to grind' > PNJ **ŋə̃*/**ŋə̃-ñ*, PSJ **ŋə̃*/**ŋə̃-ŋ*

(20) Proto-Jê **ỹ* > Proto-Cerrado **ə̃* (**ỹ* after palatals), Proto-Southern Jê **ĩ*
  a. PJ **ñỹ* 'to sit.SG' > PCerr **ñỹ/*ñỹp-r*, PSJ **nĩ* (vb. **nĩ-m*)
  b. PJ **krỹñ* 'head' > PCerr **krə̃j'*, PSJ **krĩ* (vb. **krĩ-n*)
  c. PJ **jVkrỹñ* 'knee' > PCJ **hikrə̃j* // **hikrə̃*, PSJ **jəkrĩ*

(21) Proto-Jê **õ* > Proto-Cerrado **õ*, Proto-Southern Jê **ə̃*
  a. PJ **põ* 'to rub, to clean' > PNJ **põ/*põ-ñ*, PSJ **pə̃/*pə̃-ŋ* 'to clean a field'
  b. PJ **nõ* 'to lie.SG' > PCerr **nõ/*nõp-r*, PSJ **nə̃*

(22) Proto-Jê **ũ* > Proto-Cerrado **õ*, Proto-Southern Jê **ũ*
  a. PJ **prũ* 'wife' > PCerr **prõ*, PSJ **prũ* (vb. **prũ-ŋ*)
  b. PJ **mũ* 'to go.PL' > PCerr **mõ/*mõ-r'*, PSJ **mũ* (vb. **mũ-n*)
  c. PJ **tũ* 'NEG' > PCJ **tõ*, PSJ **tũ/*tũ-ŋ* (vb. **tũ-ŋ*)
  d. PJ **ũt/*ñũt* 'to sleep' > PCerr **ŋõt/*ñõt*, PSJ **nũr*
  e. PJ **ŋũ* 'to push against, to crumble' > PNJ **ka-ŋõ/*ka-ŋõ-ñ*, PSJ **ŋũ/*ŋũ-ŋ*
  f. PJ **ŋrũ (~ *ŋrũn)* 'toucan' > PNJ **ŋrõ*, PSJ **ŋrũ*
  g. PJ **ñũctə* 'tongue' > PCerr **ñõjto*, PSJ **nũnã*
  h. PJ **ñũmjê(C)* 'female breast' > PNJ **ñõmjê*, PSJ **nũŋjê* 'female breast, milk'

(23) Proto-Jê **ẽ* > Proto-Cerrado **ẽ*, Proto-Southern Jê **ĩ*
  a. PJ **mẽ* 'PL; with' > PCerr **mẽ*, PSJ **mĩ* 'PL'
  b. PJ **tẽ* 'to go.SG' > PCerr **tẽ/*tẽm'*, PSJ **tĩ/*tĩ-ŋ* (vb. **tĩ-n*)
  c. PJ **rẽ* 'to throw.PL' > PNJ **rẽ/*rẽ-ñ*, PSJ **rĩ/*rĩ-ŋ/*rĩ-ñ*

(24) Proto-Jê **ĩ* > Proto-Cerrado **ĩ*, Proto-Southern Jê **ĩ*
  a. PJ **pĩm* 'tree, wood' > PCerr **pĩm*, PSJ **pĩ* 'fire, firewood'
  b. PJ **-cĩ(C)* 'bandage, sling' > PNJ **ja-cĩ*, **ka-cĩ*, PSJ **-θĩ* (vb. *-θĩ-n*)
  c. PJ **ñĩ* 'meat' > PCerr **ñĩ*, PSJ **nĩ*
  d. PJ **ñĩ(m)-* 'hand (*in compounds*)' > PCerr **ñĩ(m)-kra* 'hand', PSJ **nĩ-* (as in **nĩ-pê* 'to wash hands')
  e. PJ **ŋrĩ(C)* > PNJ **ku-ŋrĩ* 'to gather in a bundle', **-ŋrĩ* 'to make packages', PSJ **ŋrĩ/*ŋrĩ-ŋ*, **ŋrĩ-r* (vb. **ŋrĩ-n*) 'wrap'
  f. PJ **ñĩja* 'nose' > PCerr **ñĩja-kre*, PSJ **nĩjã*
  g. PJ **ñĩjə̂* 'smoke' > PCJ **ñĩje*, PSJ **nĩjə* (vb. **nĩja-ŋ*)

In non-final (unstressed) syllables, divergent correspondences may be observed. One such correspondence involves the PCerr formative that is reflected in PNJ as **py-/*pu-* word-initially (harmonizing in roundness with the nucleus of the root, cf. Salanova 2011a: 66) or as **-p-* word-medially, and in PCJ as **pi-*. In some stems, this formative corresponds to PSJ **pa-* (25a–b); in others, it corresponds to PSJ **pe-* (25c–d), thus evidencing a merger that would have occurred historically in PCerr. I assume that PSJ displays here the same reflexes of PJ vowels that are found in stressed syllables, whereas in the history of PCerr the unstressed vowels **o* and **e* were reduced to zero (probably via **u*/**i*, since high vowels are more likely to be syncopated). The consonant clusters created by this development were subsequently undone via different processes in PNJ (resyllabification if available, epenthesis of **y*/**u* elsewhere) and in PCJ (epenthesis of **i*).

(25) Proto-Jê **o, *e* > PCerr zero in unstressed syllables
  a. PJ **pondʊ* 'bad *or* crooked' > PCerr **pⁿndu* (PNJ **pundu* 'bad'), PSJ **pandô* (vb. **pando-n*) 'crooked'
  b. PJ **potu* 'horsefly' > PCerr **pⁿtu* (PNJ **pucu*, PCJ **pidu*), PSJ **patu*

c. PJ *peju* ‘to hide’ > PCerr **pʲju/*pʲju-r*’ (PNJ **puĵu/*puĵu-r*), PSJ **peju*
d. PJ **peti* ‘to dream’ > PCerr **pʲti/*pʲti-r*’ (PNJ **pyti/*pyti-r*), PSJ **peti/*peti-ŋ*

There appear to be no counterexamples to the development exemplified in (25): the usual reflexes of PJ **o* and **e* (PCerr **wa* > PNJ **wa, *wô*, PCJ **wa*; PCerr **e* > PNJ **e*, PCJ **ê*) do not occur in prefixes or in the unstressed syllables of underived stems in the Cerrado languages.

Only one well-known cognate set violates the regularities outlined in this section: as shown in (26), it features a correspondence between PCerr **ô* and PSJ **e*. Judging by the external cognates, such as Maxakalí *-cyc* (orthographically <*-xux*>) or Krenák *jət*, PCerr is conservative here: the correspondence between Maxakalí *y*, Krenák *ə*, and PCerr **ô* is well-attested (these vowels usually correspond to PSJ **ə*), whereas PSJ **e* usually has entirely different correspondences: Maxakalí *e*, Krenák *i*, PCerr **e* (Nikulin & Silva, forthc.).

(26) An exceptional correspondence between PCerr **ô* and PSJ **e*
PJ **côj* ‘leaf’ > PCerr **côj*’, PSJ **θej* (vb. **θe-ñ*)

I hypothesize that PSJ **e* in **θej* is a result of fronting of pre-PSJ **ə* (the regular outcome of PJ **ô*) before a palatal coda. The development **-əj* > **-ej* may have been regular: the rhyme **-əj* does not occur abundantly in Proto-Southern Jê lexicon, the only known example being **wəj* ‘dawn’ without a known external etymology. However, in the absence of supporting examples the hypothesis in question remains rather speculative.

3.3. Codas

The correspondences between the codas of PCerr and PSJ, as well as the respective Proto-Jê reconstructions are presented in Table 7 below. I remind the reader that the apostrophe stands for a suppressed echo vowel (word-finally). For Southern Jê, in addition to the codas found in underived stems, I provide information regarding the allomorphy of the verbalizer/causative suffix. Four codas (/*p’, *t’, *c’, *k’/) are reconstructed only for the underlying level of PJ.

Cognate sets that instantiate the aforementioned sound correspondences (excepts those going back to PJ underlying codas, for which see below) are provided below in (27–40).

(27) Proto-Jê **p* > Proto-Cerrado **p*, Proto-Southern Jê **∅*
a. PJ **kôp* ‘fly, mosquito’ > PCerr **kôp* ‘fly’, PSJ **kə* ‘mosquito’

(28) Proto-Jê **t* > Proto-Cerrado **t*, Proto-Southern Jê **r* (vb. **-n*)
a. PJ **mbyt* > PCerr **mbyt* ‘sun’, Ingain *pyr* ‘moon’
b. PJ **tât* ‘strong, hard’ > PCerr **tât*, PSJ **tər* (vb. **tan*)
c. PJ **rĩt* ‘to look’ > PCerr **rĩt*, PSJ **rĩr*
d. PJ **ñũt* ‘to sleep.NF’ > PCerr **ñõt*, PSJ **nũr*
e. PJ **kẽt* ‘stone’ > PCerr **kẽt*, Ingain *kẽr*
f. PJ **krot* ‘chin, beak’ > PCerr **krwat* ‘beak’, Ingain *jat-krar* ‘chin’

(29) Proto-Jê **c* > Proto-Cerrado **c*, Proto-Southern Jê **∅* (vb. **-n*)
a. PJ **pôc* ‘to leave.PL’ > PCerr **pôc*, PSJ **pə* (vb. **pə-n*)
b. PJ **mbec* ‘good’ > PCerr **mbec*, PSJ **mbe* ‘HABIT’, **ã-mbe* (vb. **ã-mba-n*) ‘good weather’
c. PJ **ñũctə* ‘tongue’ > PCerr **ñõjto*, PSJ **nũnã*

(30) Proto-Jê **k* > Proto-Cerrado **k*, Proto-Southern Jê **∅* (vb. **-n*) or **j* (after **ê*)
a. PJ **tyk* ‘black, dark’ > PCerr **tyk* ‘black’, PSJ **ku-ty* (irreg. vb. **ku-ty-ŋ*) ‘dark, night’
b. PJ **kôk ~ *ka-kôk* ‘wind’ > PCerr **kôk, *ka-kôk*, PSJ **kã-kə* (vb. **kã-kə-n*)

Table 7. Proto-Jê codas and their reflexes in Proto-Cerrado and Proto-Southern Jê.

PJ	PCerr	PSJ	PSJ (vb.)
followed by echo vowels			
*p	*p	*ø	*-n
*t	*t	*r	
*c	*c	*ø	
*k	*k	*ø, *j†	
*m	*m	*ø	
*n	*n		
*ñ	*j’		
*r	*r	*n	*-n
*j	*j’	*j	*-ñ

PJ		PCerr	PSJ	PSJ (vb.)
no echo vowels				
*ø	*/p’/	*ø, *p‡	*ø	*-m
	*/t’/	*ø		*-n
	*/c’/			*-ñ
	*/k’/			*-ŋ
*m’		*m’	*ŋ	*-ŋ
*n’		*n’	*n	*-n
*ñ’		*ñ’	*ñ	*-ñ
*ŋ’		(?)	*ŋ	*-ŋ
*r’		*d	*n	*-n
*j’		*c	*j	*-ñ

† after *ê[17] (no vb. attested, but probably *ñ); ‡ in non-finite forms preceding the suffix *-r

c. PJ *pêk ‘to fart.NF’ > PNJ *pê-k, PSJ *pêj (but PCJ *pi-ri)
d. PJ *kujêk ‘vein’ > PNJ *kuĵêk, PSJ *kujêj

(31) Proto-Jê *m > Proto-Cerrado *m, Proto-Southern Jê *ø (vb. *-n)
a. PJ *pĩm ‘tree, wood’ > PCerr *pĩm, PSJ *pĩ ‘fire, firewood’
b. PJ *ndəm ‘eye’ > PCerr *ndom, PSJ *kə-nã (vb. *kə-nã-n) ‘eye, fruit’
c. PJ *cym ‘seed’ > PCerr *cym, PSJ *θy (vb. *θy-n)
d. PJ *kôm ‘tree, horn’ > PCerr *kôm, PSJ *kə ‘tree’, *nĩ-kə ‘horn’

(32) Proto-Jê *n > Proto-Cerrado *m, Proto-Southern Jê *ø (vb. *-n)
a. PJ *mbyn ‘tail’ > PCerr *mbyn, PSJ *mby
b. PJ *mbên ‘liquid’ > PCerr *mbên, PSJ *mbê (vb. *mbe-n)

(33) Proto-Jê *ñ > Proto-Cerrado *j’, Proto-Southern Jê *ø (vb. *-n)
a. PJ *pañ ‘arm, branch’ > PCerr *paj’, PSJ *pa
b. PJ *jañ ‘to eat.INTR, food’ > PCJ *caj // *ca, PSJ *jã/*jã-ŋ/*jã-n (vb. *jã-n)
c. PJ *krỹñ ‘head’ > PCerr *krə̃j’, PSJ *krĩ (vb. *krĩ-n)
d. PJ *krıñ ‘thigh’ > PCerr *k(r)jaj’, PSJ *krê

(34) Proto-Jê *r > Proto-Cerrado *r, Proto-Southern Jê *n
a. PJ *par ‘foot’ > PCerr *par, PSJ *pãn

[17] I thank Mário André Coelho da Silva for pointing out the possibility that PJ *-êk > PSJ *-êj could be a regular sound change.

(35) Proto-Jê **j* > Proto-Cerrado **j'*, Proto-Southern Jê **j* (vb. **-ñ*)
  a. PJ **côj* 'leaf' > PCerr **côj'*, PSJ **θej* (vb. **θe-ñ*)

(36) Proto-Jê **m'* > Proto-Cerrado **m'*, Proto-Southern Jê **ŋ*
  a. PJ **prə̃m'* 'hungry' > PCerr **prə̃m'*, PSJ **prə̃ŋ* 'hungry; year'
  b. PJ **mbrʊm'* 'ant' > PNJ **mbrũm*, PSJ **mbroŋ-jy* 'Argentine ant'
  c. PJ **k(r)Vmtym'* 'capybara' > PCerr **kumtym'*, PSJ **kryŋndyŋ*
  d. PJ **tom'* 'fat' > PCerr **twam'*, PSJ **taŋ*
  e. PJ **tʊm'* 'old' > PCerr **tum'* 'old, tall'; PSJ **toŋ* 'dry (*of plants*)'
  f. PJ **ñũmjê(C)* 'female breast' > PNJ **ñõmjê*, PSJ **nũŋjê* 'female breast, milk'

(37) Proto-Jê **n'* > Proto-Cerrado **n'*, Proto-Southern Jê **n*
  a. PJ **mbɪn'* 'husband' > PNJ **mbjên*, PSJ **mben*

(38) Proto-Jê **ñ'* > Proto-Cerrado **ñ'*, Proto-Southern Jê **ñ*
  a. PJ **ŋgoñ'* 'feather, hair' > PNJ **ŋgwə̂ñ* 'feather', PSJ **ŋgañ* 'animal hair'

(39) Proto-Jê **r'* > Proto-Cerrado **d*, Proto-Southern Jê **n*
  a. PJ **jar-ko* 'mouth' > PCerr **jad-kwa*, PSJ **jãn-ka* 'door'

(40) Proto-Jê **j'* > Proto-Cerrado **j'*, Proto-Southern Jê **j* (vb. **-ñ*)
  a. PJ **kaj'* 'basket' > PNJ **kac*, PSJ **kãj* (vb. **kãñ*)

The reconstruction of PJ **ŋ'* is supported by only one PSJ token, PSJ **mə̃ŋ* 'honey bee', which lacks known cognates in Cerrado languages but corresponds externally to Maxakalí *pyk* (orthographically <*puk*>) and Krenák *pəŋ*. As Nikulin and Silva (forthc.) show, the point of articulation of Proto-Jê codas systematically matches that of Maxakalí, and the manner of articulation of Proto-Jê codas systematically matches that of Krenák.

My reconstruction of Proto-Jê underlying codas (/*p', *t', *c', *k'/) is crucially based on morphophonological evidence from Central and Southern Jê and is corroborated by external comparison with Maxakalí and Krenák (Nikulin & Silva, forthc.). No contemporary Jê language shows any reflex of these codas in underived environments, which probably means that they surfaced as zero in underived Proto-Jê words. In Central Jê, only PJ /*p'/ left any trace: in non-finite forms of verbs, it combined with the non-finite suffix **-r*, yielding PCJ **-b-rV*. In Southern Jê, the quality of the underlying codas is made manifest in derived denominal verbs or causatives, where underlying /-n/ acquires the point of articulation of the underlying coda consonant (see Cavalcante 1987: 51–52 for a synchronic analysis for São Paulo Kaingáng). Examples are given in (41–44).

(41) Proto-Jê */p'/ > Proto-Cerrado *∅ (NF **-p-r*), Proto-Southern Jê *∅ (vb. **-m*)
  a. PJ **tə* */təp'/ 'to fly.SG' > PCerr **to*/**top-r*, PSJ **tã* (vb. **tã-m*)
    cf. Maxakalí /tup-a/, IRR /tup/ 'id.'
  b. PJ **nõ* */nõp'/ 'to lie.SG' > PCerr **nõ*/**nõp-r*, PSJ **nə̃*
    cf. Maxakalí /nũp/ 'to lie.PL'
  c. PJ **cy* */cyp'/ 'to weave' > PCerr **cy̲*/**cy̲p-r'*, PSJ **θy*
    cf. Maxakalí /cap/ 'id.'
  d. PJ **cô* */côp'/ 'to eat soft food, to suck' > PNJ **cô*/**cô-r*, PSJ **pe-θə* (vb. **pe-θa-m*) 'to suck breast'
    cf. Maxakalí /cyp/ 'to suck, to lick'
  e. PJ **nj̲ô* */ñôp'/ 'to hang.SG' > PCerr **nj̲ô*/**nj̲ôp-r'*, PSJ **cə*/**ca-ŋ* (vb. **cə-m*)
    cf. Maxakalí /cyp/ 'id.'

f. PJ **ñỹ* */jỹp'/ 'to sit.SG' > PCerr **ñỹ/*ñỹp-r*, PSJ **nĩ* (vb. **nĩ-m*)
cf. Maxakalí /ñỹp/ 'id.'
g. PJ **jʊ* */jʊp'/ 'pus' > PCerr **jup-r*, PSJ **θô* (vb. **θo-m*)
h. PJ **ũ ~ *õ* /*ũp' ~ *õp'/ 'to give' > PCerr **ŋõ/*ñõp-r*'
cf. Maxakalí /hũp/ 'id.'

(42) Proto-Jê */t'/ > Proto-Cerrado *∅, Proto-Southern Jê *∅ (vb. **-n*)
a. PJ **jarê* (~ **jarên*) */jarêt' ~ jarên/ 'root' > PNJ **jarê*, PSJ **jãrê* (vb. **θ-are-n*)
cf. Maxakalí /ñĩp-catit/ 'id.'
b. PJ **jê* (~ **jên*) */jêt' ~ jên/ 'thread, knot, cloth' > PNJ **jê*, PSJ vb. **θe-n* 'to spin thread'
cf. Maxakalí /cit/ 'id.'
c. PJ **jiji ~ *ñĩji* */jijit' ~ jĩjit'/ 'name' > PCerr **ñĩji*, PSJ **jiji ~ *jyjy* (vb. **jiji-n ~ *jyjy-n*)
cf. Maxakalí /-cit/ 'id.'
d. PJ **ko* */kot'/ 'to dig' > PNJ **kwô/*kwô-ñ*, PSJ **ka/*ka-ŋ* (vb. **ka-n*)
cf. Maxakalí /kut/ 'id.'
e. PJ **ŋgô* */ŋôt'/ 'louse' > PCerr **ŋgô*, PSJ **ŋgə* (vb. **ŋga-n*) 'louse, flea'
cf. Maxakalí /kyt/ 'id.'

(43) Proto-Jê */c'/ > Proto-Cerrado *∅, Proto-Southern Jê *∅ (vb. **-ñ*)
a. PJ **prʏ* */prʏc'/ 'coal, ember' > PCerr **prə* 'ember', PSJ **prô-ñ*[18]
b. PJ **jʏ* */jʏc'/ 'urine' > PCerr **jə* 'urine', PSJ vb. **jô-ñ* 'to urinate'
cf. Maxakalí /cyc/ 'id.'
c. PJ **jacê ~ *jVmcê* 'nest' */jacêc' ~ jamcêc'/ > PCerr **jacê*, PSJ **jəŋθê* (vb. **jəŋθe-ñ*)

(44) Proto-Jê */k'/ > Proto-Cerrado *∅, Proto-Southern Jê *∅ (vb. **-n*)
a. PJ **tũ* */tũk'/ 'NEG' > PCJ **tõ*, PSJ **tũ/*tũ-ŋ* (vb. **tũ-ŋ*)
cf. Maxakalí /nũk/ 'to end, to run out'
b. PJ **kra* */krak'/ 'offspring' > PCerr **kra*, PSJ **krã* (vb. **krã-ŋ*)
cf. Maxakalí /ktuk/ 'id.'

I do not represent these underlying codas in my reconstructions throughout this paper; the examples (41–44) above include almost every cognate set for which the presence of an underlying coda may be established with any certainty.

4. Jê etymologies

In this section, we present the most reliable cognate sets identified so far. The data are sorted by the onset of the final (stressed) syllable, then by its nucleus, then by its coda, then by any preceding material, following the order /p, pr, m, mr, w, t, n, d, r, c, ñ, j, k, kr, ŋ, ŋr, a, ə, ə̃, ə̂, ʏ, y, ỹ, o, õ, ô, ʊ, u, ũ, e, ẽ, ê, ɪ, i, ĩ/. I list first the etymologies that have reflexes both in Cerrado and in Southern Jê languages, and then proceed to the reconstructions that are based only on Cerrado or Southern Jê reflexes (as well as their external cognates elsewhere in Macro-Jê or Tupian[19]), respectively.

[18] It is unclear why the PSJ reflex looks like a verbalized form.

[19] Although I do not currently regard the relation between Macro-Jê and Tupian as conclusively proven, I still find the hypothesis quite promising. Some novel lexical evidence may be found in Nikulin & Silva forthc. (as well as in this section). The Proto-Tupian reconstructions are my own and follow the principles outlined by Carvalho and Nikulin (in prep.).

For reasons of space, I do not provide the data of individual Northern Jê languages, limiting myself to Proto-Northern Jê reconstructions. Detailed information on the reflexes of Proto-Northern Jê items in individual Northern Jê languages may be found in Nikulin & Salanova (forthc.). Ingain and Southern Kayapó data are only given for cognate sets that lack known reflexes in Kaingáng/Laklãnõ and Panará, respectively.

Reflexes in both branches

PJ **par* 'foot':
PCerr **par* > PNJ **par*; PNR *pa:*; PCJ **para* // **pa:ra* (XAV *para* // *pa:ra*, XER *pra*)
PSJ **pãn* > KGG *pẽn*; LKL *pãn*

PJ **pañ* 'arm, branch':
PCerr **paj'* > PNJ **pa*; PNR *pa*; PCJ **paj-nõ* 'arm' (XAV *pan-nõ*, XER *paj-nõ*), **paj-hi* (XAV *paj-hi*); (?) PCJ **pa-krata* // **pa-kra:da* 'root' (XER *pa-krta* // *pa-krda*), (?) **wêdê-pa* 'root' (XAV *wêdê-pa*, XER *wdê-pa*)
PSJ **pã* > KGG *pẽ*; LKL *pã*

PJ **pĩm* 'tree, wood, firewood':
PCerr **pĩm* > PNJ **pĩ*; PNR *pĩ*; PCJ **mĩmĩ* // **mĩ* (XAV *mĩmĩ* // *mĩ*, XER *mmĩ*)
PSJ **pĩ* 'fire, firewood' > KGG *pĩ*; LKL *pẽ*

PJ **põ* 'to rub, to clean':
PCerr **põ*/**põ-ñ'* > PNJ **põ*/**põ-ñ*
PSJ **pə̃*/**pə̃-ŋ* 'to clean a field' > KGG *pə̃*/*pə̃-ŋ*

PJ **pôc* 'to leave.PL':
PCerr **pôc* > PNJ **to=pôj*/**to=pôc* 'to extract.PL'; PCJ **pu*/**pu-ci* 'to leave.DU' (XAV *pu-ci*, XER *pu*/*pu-s(i)*)
PSJ **pə* (vb. **pə-n*) > KGG *pa* (vb. *pa-n*); LKL *po*

PJ **pê*/**pê-k* 'to fart':
PCerr **pê*/**pê-k* > PNJ **ij-pê*/**pê-k*; PCJ **pi*/**pi-ri*[20] (XAV *pi*/*pi-ri*)
PSJ **pêj* > KGG *pêj*

PJ **prə̃'* 'hungry, to want':
PCerr **prə̃m'* > PNJ **prə̃m*; PNR *pjə̃nĭ* 'to want'; PCJ **mrə̃m* // **mrə̃* 'hungry' (XAV, XER *mrə̃m* // *mrə̃*)
PSJ **prə̃ŋ* 'hungry; year' (probably via 'hunger season') > KGG *prə̃ŋ*; LKL *plõŋ* 'year'

PJ **prʏ* 'coal, ember':
PCerr **prə* 'ember' > PNJ **prə*; PCJ **pre* (XER *pre-hika* 'glowing embers', *pre-nĩzuri* 'spark', *pre-zapdo* 'ember')
PSJ **prə̂ñ* > KGG *prəñ*; LKL *pləñ*

PJ **pry (~ *pryn)* 'road':
PCerr **pry (~ *pryn)* > PNJ **pry*; PNR *pjy*
PSJ **ã-pry*/**NP *jə-pry* > KGG *ẽ-pry*/NP *ja-pry*

PJ **prũ* 'wife':
PCerr **prõ* > PNJ **prõ*; PCJ **mrõ* 'spouse, to marry' (XAV, XER *mrõ*)
PSJ **prũ* (vb. **prũ-ŋ*) > KGG *prũ* (vb. *prũ-ŋ*); LKL *plũ* (vb. *plũ-ŋ*)

PJ **mba* 'to hear, to understand':
PCerr **mba*/**mba-r'* > PNJ **mba*/**mba-r*; PNR *impa-rĭ*; PCJ **wa-pa*/**wa-pa-ri* (XAV *wa-pa*/*wa-pa-ri*, XER *wa-pa*/*wa-pa-r(i)*)
PSJ **mə̃*/**mə̃-ŋ* > KGG *mẽ*/*mẽ-ŋ*; LKL *mə̃*/*mə̃-ŋ*

[20] The expected non-finite form would be ***pi-ki* // ***pi*. Northern Jê is considered to be more conservative than Central Jê here because the Southern Jê cognate appears to correspond to the Northern Jê non-finite form.

PJ **mba* ‘liver’:
PCerr **mba* > PNJ **mba*; PNR *impa*; PCJ **pa* (XAV, XER *pa*)
PSJ **tə̃-mã* > KGG *tə̃-mẽ*; LKL *tõ-mã*
PJ **-mbaŋ’*[21] ‘to be afraid’:
PCerr **pᵊ-mba* > PNJ **pymba*; PNR *s-umpa*; PCJ **pipa* (XAV, XER *pipa*)
PSJ **ka-mãŋ*, **mũ-mãŋ* > KGG *ka-mẽŋ, mũ-mẽŋ*; LKL *ko-mãŋ, mõ-mãŋ*
PJ **mə̃* ‘DAT’:
PCerr **mə̃* > PNJ **mə̃*; PNR *mə̃*; PCJ **mə̃* (XAV, XER *mə̃*)
PSJ **mə̃* > KGG *mə̃*; LKL *mõ*
PJ **mbə̂* ‘to grab, to carry’:
PCerr **mbə̂/*mbə̂-ñ* > PNJ **mbə̂/*mbə̂-ñ* ‘to grab’; PNR *impə-rĭ* ‘to carry’; PCJ **kwa-pe/*kwa-pe-j // *kwa-pe* ‘to carry.DU’ (XAV *ʔwa-pe/ʔwa-pe-j // ʔwa-pe*, XER *kwape*)
PSJ **mbə/*mba-ŋ, *mbə-r* (vb. **mba-n*) ‘to hold, to carry (a short object)’ > KGG *mba/mõ-ŋ/mõ-n*’; LKL *mbo/mba-ŋ, mbol* (vb. *mba-n*)
PJ **mbyn* ‘tail’:
PCerr **mbyn* ‘tail, penis’ > PNJ **mby* ‘penis’, **ja-mby* ‘tail’; PNR *s-ampy* ‘tail’; PCJ **mə̃nə̃ // *bə* (XAV *mə̃nə̃ // bə*, XER *mnə̃ // bə*)
PSJ **mby* > KGG, LKL *mby*
PJ **mbyt* ‘celestial body’:
PCerr **mbyt* ‘sun’ > PNJ **mbyt*; PNR *impytĭ* ‘time’; PCJ **bətə // *bɔ̈də* (XAV *bətə // bɔ̈də*, XER *btə // bdə*)
Ingain[22] *pyr* ‘moon’: *Lista* ⟨pirihí⟩, *Pedro, Cosme Román* ⟨puirí⟩, *María Antonia* ⟨puiré⟩, *Vogt*[1] ⟨pyri⟩, *Vogt*[3] ⟨pyrý⟩
PJ **mbec* ‘good’:
PCerr **mbec* > PNJ **mbec*; PNR *impe* ‘real’; PCJ **pêcê // *pê* (XAV *pêcê // pê* ‘well; to recover, to get better’, XER *pêsê // pê*)
PSJ **mbe* ‘HABIT’ > KGG, LKL *mbe*; PSJ **ã-mbe* (vb. **ã-mba-n*) ‘good weather’ > KGG *ẽ-mbe* (vb. *ẽ-mõ-n*)
PJ **mẽ* ‘PL; with’:
PCerr **mẽ* > PNJ **mẽ*; PNR *-mẽ-ra* ‘PL (nominal number)’, *mẽ=* ‘DU (indexed on verbs)’; PCJ **mẽ* ‘with’ (XAV *mẽ*, XER *mẽ*)
PSJ **mĩ* ‘PL’ > LKL *mẽ*
PJ **mbên* ‘liquid’:
PCerr **mbên* > PNJ **mbê*; PCJ **pĩnĩ // *pi* (XAV *pĩnĩ // pĩ*[23] ‘honey’, XER *to-pi* ‘eye rheum’)
PSJ **mbê* (vb. **mbe-n*) > KGG *mbê* (vb. *mbe-n*); LKL *mbê*
PJ **mbɪn’* ‘husband’:
PCerr **mbjan’* > PNJ **mbjên*; PNR *impin-pjə*[24]
PSJ **mben* > KGG, LKL *mben*

[21] PSJ **-ŋ* in underived stems usually goes back to PJ **-m’* and corresponds to PCerr **-m’*. The correspondence between PSJ **-ŋ* and PCerr zero is unparalleled; I provisionally reconstruct PJ **-ŋ’*.

[22] Hereinafter, *Lista* refers to data from Lista 1883; *Pedro, Cosme Román,* and *María Antonia* are the names of Ambrosetti’s (1896) consultants who provided the respective pieces of data; *Vogt*[1]*, Vogt*[3] refer to two original word-lists contained in Vogt 1904.

[23] The expected utterance-final allomorph would be ***pi*. One possible explanation for this irregular reflex is an analogical extension of nasality from the regular utterance-medial allomorph, *pĩnĩ*.

[24] It is unclear whether the Panará reflex of PCerr **-ja-* is regular.

PJ **mũ* 'to go.PL':
PCerr **mõ/*mõ-r'* > PNJ **mõ/*mõ-r*; PNR *mõ-rĭ*; PCJ **mõ/*mõ-rĩ* 'to go.SG' (XAV *mõ/mõ-rĩ*, XER *mõ/mõ-r(ĩ)*)
PSJ **mũ* (vb. **mũ-n*) > KGG *mũ* (vb. *mũ-n*); LKL *mũ*
PJ **mbrə* 'ashes':
PCerr **mbro* > PNJ **mbro*; PCJ **wêdê-pro* 'coal' (XAV *wêdê-pro*, XER *wdê-pro*), **kə ĵaj-mpro* 'foam' (XAV *ʔə ĵaj-pro*, XER *kə zaĩ-pro ~ kə zam-pro*), **ĵadaj-mpro* 'saliva' (XAV *ĵadaj-pro*, XER *zdaĩ-pro*), (?) **pro* 'to burn' (XER *pro*)
PSJ **mrã* > KGG *mrẽj*;[25] LKL *mlã*
PJ **mbrʊm'* 'ant':
PCerr **mbrum'* > PNJ **mbrũm*
PSJ **mbroŋ-jy* 'Argentine ant' > KGG *mbroŋ-jy*
PJ **mbre* 'relative by marriage' (*kinship term*):
PCerr **mbre* > PNJ **mbre* 'sibling-in-law'; PCJ **mə-prê-baba // *mə-prê-wa* (XAV *mə-prê-baba // mə-prê-wa* 'mother/father-in-law', XER *mə-prê-wa*)
PSJ **jə-mbre* > KGG *ja-mbre* 'cousin, son-in-law'; LKL *jo-mble* 'brother-in-law, father-in-law, mother-in-law'
PJ **wa* 'to walk':
PCerr **wa* > PNJ **ba* 'to walk, to live'; PNR *pa*
PSJ **wa* 'STAT' (aspect marker) > KGG *wẽ*; LKL *vã*
PJ **wy* 'to take, to carry':
PCerr **wy/*wy-r'* > PNJ **by/*by-r*; PNR *py-rĭ*
PSJ **wə/*wa-ŋ/*wã-n ~ *wy-n*[26] 'to hold, to carry (a long object)' > KGG *wa/wõ-ŋ/wõ-n ~ wy-n*; LKL *va/võ-ŋ/võ-n ~ vu-n*
PJ **wô* 'to untie':
PCerr **wô/*wô-ñ'* > PNJ **bô/*bô-ñ*; Southern Kayapó ⟨tipó⟩ (likely *pô*)
PSJ **kə-wə/*kə-wa-ŋ, *kə-wə-r* (vb. **kə-wa-n*) 'to let loose, to untie' > KGG *ka-wa/ka-wõ-ŋ, ka-wa-r* (vb. *ka-wõ-n*)
PJ **tə* 'to fly.SG':
PCerr **to/*top-r* > PNJ **to/*to-r*; PNR *to:/to:-j* 'to fly, to dance'; PCJ **tob-ro* (XAV *tob-ro* 'to descend', XER *tb-ro* 'to cross a body of water')
PSJ **tã* (vb. **tã-m*) > KGG *tẽ* (vb. *tẽ-m*); LKL *tã* 'to approach.SG'
PJ **ñũctə* 'tongue':
PCerr **ñõjto* > PNJ **ñõto*; PNR *s-õto*; PCJ **ñõjto* (XAV *ñõjto ~ ñõtto*, XER *nõjto*)
PSJ **nũnã* > KGG *nũnẽ*; LKL *nũnã*
PJ **tə̃* 'INSTR'
PCerr **tə̃* > PNJ **tə̃* 'LOC'; PNR *tə̃* 'ALL' or *hə̃/(r)ahã* 'ADESS'; PCJ **nə̃* (XAV, XER *nə̃*)
PSJ **tə̃* 'ERG, INSTR' > KGG *tə̃*; LKL *tõ*
PJ **tât* 'strong, hard':
PCerr **tât* > PNJ **tâc*; PNR *tətĭ*; PCJ **têtê // *tê:dê* (XAV *-têtê // tê:dê*, XER *têt(ê) ~ ttê, -têt(ê) ~ -ttê // -tdê*)
PSJ **tər* (vb. **ta-n*) > KGG *tar* (vb. *tõ-n*); LKL *tol*
PJ **ty* 'to die':

[25] The accretion of Kaingáng **-j* is irregular.

[26] The relation of PSJ **wy-n* to the remaining forms is not a regular one; it might be the case that two distinct etymologies are conflated here. Moreover, the Laklãnõ reflex unexpectedly shows a rounded vowel.

PCerr **ty(r)/*ty-k*[27] > PNJ **ty/*ty-k*; PNR *ty*; PCJ **dərə/*də-kə // *də* (XAV *dərə/də-ʔə // də*, XER *dərə/d-kə // də*)

PSJ **ty* 'to die.STAT' > KGG *ty* 'numb'; LKL *ty*

PJ **k(r)Vmtym'* 'capybara':

PCerr **kumtym'* > PNJ **kũmtũm*;[28] PNR *intyŋ*; PCJ **kumdəm // *kumdə* (XAV *ʔumdə*, XER *kumdəm // kumdə*)

PSJ **kryŋndyŋ* > KGG *kryŋndyŋ*; (?) LKL *klẽ̂juñ*

PJ **tyk* 'black, dark':

PCerr **tyk* 'black' > PNJ **tyk*; PNR *kə-ty:*; PCJ **dəkə // *də* (XAV *ʔrə̃-də*, XER *dkə* 'dark')

PSJ **ku-ty* (vb. **ku-ty-ŋ*) 'dark, night' > KGG, LKL *ku-ty* (vb. *ku-ty-ŋ*)

PJ **tom'* 'fat':

PCerr **twam'* > PNJ **twỗm*; (?) PNR *tumə̃*; PCJ **wam // *wa* (XAV *wam // wa*, XER *rom-wa*)

PSJ **taŋ* > KGG *tõŋ*; LKL *taŋ*

PJ **tʊm'* 'old':

PCerr **tum'* 'old' > PNJ **tũm*; PNR *tuŋ*; PCJ **dum // *du* 'tall' (XER *dum*), **dum-krata // *dum-kra:da* 'elder sibling of the opposite sex' (XAV *dub-ʔrata // dub-ʔra:da*, XER *dum-krda*)

PSJ **toŋ* 'dry (*of plants*)' > KGG *toŋ*; LKL *tuŋ*

PJ **tu* 'to carry':

PCerr **tu/*tu-r'* > PNJ **tu/*tu-r*; PNR *tu-rĩ* 'to carry in a basket.SG'; PCJ **du/*du-ri* 'to carry.SG' (XAV *du/du-ri*, XER *du/du-r(i)*)

PSJ **tu* 'to carry on one's back, to wear' > KGG, LKL *tu*

PJ **potu* 'horsefly':

PCerr **pᵊtu* > PNJ **pucu*;[29] PNR *pu:su*; PCJ **pidu* (XAV, XER *pidu*)

PSJ **patu* > KGG *põtu*; LKL *patu*

PJ **tũ* 'NEG':

PCerr **tõ* > (?) PNR *rõ* 'privative; negation in non-finite clauses'; PCJ **tõ*[30] (XAV *tõ* 'prohibitive; negation in purpose clauses', XER *tõ* 'privative')

PSJ **tũ/*tũ-ŋ* (vb. **tũ-ŋ*) > KGG *tũ/tũ-ŋ* (vb. *tũ-ŋ*); LKL *tũ/tũ-ŋ*

PJ **tẽ* 'to go.SG':

PCerr **tẽ/*tẽ-m'* > PNJ **tẽ/*tẽ-m*; PNR *tẽ/tẽ-ri* 'to leave, to fall'; PCJ **nẽ/*nẽ-m // *nẽ* 'to go.DU' (XAV *nẽ/nẽ-m // nẽ*, XER *nẽ/nẽ-m(ã) // nẽ*)

PSJ **tĩ/*tĩ-ŋ* (vb. **tĩ-n*) > KGG *tĩ/tĩ-ŋ* (vb. *tĩ-n*); LKL *tẽ/tẽ-ŋ* (vb. *tẽ-n*)

PJ **peti* 'to dream':

PCerr **pᵊti/*pᵊti-r'* > PNJ **pyti/*pyti-r*; Southern Kayapó ⟨iúpintín⟩ (likely *pyti-ŋ*)

PSJ **peti/*peti-ŋ* > KGG *peti/peti-ŋ*; LKL *vãñ-mbi[ŋ]ti*

PJ **jandə* 'to send':

PCerr **jando/*jando-r'* > PNJ **jando/*jando-r*; PNR *s-anto-ri*; PCJ **ĵatõ/*ĵatõ-rĩ* (XAV *ĵatõ/ĵatõ-rĩ*, XER *zatõ/zatõ-r(ĩ)*)

PSJ **jãnã/*jãnã-ŋ* > KGG *jẽnẽ/jẽnẽ-ŋ*; LKL *jãnã/jãnã-ŋ*

PJ **ndəm* 'eye':

PCerr **ndom* > PNJ **ndo(p-)*; PNR *into*; PCJ **tõmõ // *to* (XAV *tõmõ/to*, XER *tmõ/to*)

PSJ **kə-nã* (vb. **kə-nã-n*) 'eye, fruit' > KGG *ka-nẽ*; LKL *ko-nã*

---

[27] Northern Jê languages point to PCerr finite **ty*; Central Jê languages point to PCerr finite **tyr*.

[28] The rounding of PCerr **y* in PNJ is irregular.

[29] Medial **-c-* in place of the expected ***-t-* is irregular.

[30] It is unclear why PCerr **t* yields PCJ **t* in a nasal environment here. The expected reflex would be PCJ ***nõ*.

PJ **nə̃* ‘mother’:
PCerr **nə̃* > PNJ **nə̃*; PNR *nə̃-pjə*; PCJ **nə̃* (XAV *nə̃*)
PSJ **nə̃* > KGG *nə̃*; LKL *nõ*
PJ **nõ* ‘to lie.SG’:
PCerr **nõ/*nõp-r* > PNJ **nõ/*nõ-r*; PNR *nõ*; PCJ **nõm-rõ* (XAV *nõm-rõ*, XER *nõm-r(õ) ~ nm-rõ*)
PSJ **nə̃* > KGG *nə̃*; LKL *nõ*
PJ **pondʊ (~ *pondʊn)* ‘bad *or* crooked’:
PCerr **pⁿndu (~ *pⁿndun)* ‘bad’ > PNJ **pundu*
PSJ **pandô* (vb. **pando-n*) ‘crooked’ > KGG *põndô* (vb. *põndo-n*); LKL *pandô*
PJ **ra* ‘to hit’:
PCerr **-ra/*-ra-ñ’* > PNJ **ku-ra/*ku-ra-ñ*
PSJ **rã/*rã-ŋ/*rã-n* > KGG *rẽ/rẽ-ŋ/rẽ-n*; LKL *lã/lã-ŋ/lã-n*
PJ **jara* ‘wing; armpit’:
PCerr **jara* > PNJ **jara* ‘wing’, **jara-kre* ‘armpit’; PNR *s-aja* ‘wing, feather’, *s-a-kre* ‘armpit’
PSJ **θ-ãr* ‘wing’ > KGG *ɸẽr*; LKL *ð-ãr*; PSJ **jãrã* (vb. **jãrã-n*) ‘armpit’ > KGG *jẽrẽ* (vb. *jẽrẽ-n*); LKL *jãlã*
PJ **roñ* ‘celestial body’:
PCerr **rwaj’* ‘moon’ > PNJ **mbyt-rwô*; PCJ **waj // *wa* (XER *waj // wa*)
PSJ **ra* ‘sun’ > KGG *rõ*; LKL *la* ‘sun, day’
PJ **ru* ‘to transport liquid or to pour’:
PCerr **ru/*ru-ñ’* > PNJ **ru/*ru-ñ* ‘to spill, to pour’
PSJ **ru/*ru-ŋ/*ru-n* ‘to fetch water’ > KGG *ru/ru-ŋ/ru-n*; LKL *lu-n* ‘well’
PJ **re* ‘to leave, to abandon’:
PCerr **re*[31] > PNJ **re/*re-(r)*; PCJ **rẽ/*rẽ-mẽ* (XAV *rẽ/rẽ-mẽ*, XER *rẽ/rẽ-m(ẽ) ~ r-mẽ*)
PSJ **re/*ra-ŋ* > KGG *re/rõ-ŋ*
PJ **rẽ* ‘to throw.PL’
PCerr **rẽ/*rẽ-ñ’* > PNJ **rẽ/*rẽ-ñ*
PSJ **rĩ/*rĩ-ŋ/*rĩ-ñ* > KGG *rĩ/rĩ-ŋ/rĩ-ñ* ‘to carry.PL’; LKL *lẽ/lẽ-ŋ/lẽ-ñ*
PJ **jarê* (~ **jarên*) ‘root’:
PCerr **jarê* (~ **jarên*) > PNJ **jarê*; PNR *s-arê*
PSJ **jãrê* (vb. **θ-are-n*) > KGG *jõrê*; LKL *jãlê* (vb. *ð-ale-n*)
PJ **-rê(C)* ‘to weed’:
PCerr **-rê(C)/*-rê(C)-r* > PNJ **ka-rê/*ka-rê-r*
PSJ **ku-rê/*ku-re-ŋ* ‘to weed, to cut clean’ > KGG *ku-rê/ku-re-ŋ*; LKL *ku-lê/ku-le-ŋ*
PJ **rĩt* ‘to look’:
PCerr **rĩt* > PNJ **rĩt*; PCJ **rĩtĩ // *rĩːnĩ* (XAV *rĩtĩ // rĩːnĩ* ‘to look for’, XER *rĩtĩ ~ rtĩ*)
PSJ **rĩr* ‘to wake up, to be alive’ > KGG *rĩr*; LKL *lẽl*
PJ **cy* ‘to weave’:
PCerr **cy/*cyp-r’* > PNJ **cy/*cy-r*; PCJ **ñə̃/*ñə̃m-rĩ* (XAV *ñə̃/ñə̃m-rĩ*, XER *nə̃m-r(ĩ)*)
PSJ **θy* > KGG *ɸy*; LKL *ðy*
PJ **cym* ‘seed’:
PCerr **cym* > PNJ **cy*; PNR *sy*; PCJ **ñə̃mə̃ // *ĵə* (XAV *ñə̃mə̃ // ĵə*, XER *zə*)
PSJ **θy* (vb. **θy-n*) > KGG *ɸy* (vb. *ɸy-n*); LKL *ðy*
PJ **cô* ‘to eat soft food, to suck’:
PCerr **cô/*cô-r’* > PNJ **cô/*cô-r*; PNR *sôw-rĭ*
PSJ **pe-θə* (vb. **pe-θa-m*) ‘to suck breast’ > KGG *pa-ɸa*[32] (vb. *pa-ɸõ-m*); LKL *pe-ðo* (vb. *pe-ða-m*)

[31] It is unclear how the non-finite form of this verb should be reconstructed.

PJ **côj* ‘leaf’:
PCerr **côj’* > PNJ **cô*; PNR *pərĭ-sô*; PCJ **cuj // *cu* (XAV *wê-cuj-rə, -cuj // -cu*, XER *su*)
PSJ **θej* (vb. **θe-ñ*) > KGG *ɸej*; LKL *ðej*
PJ **jacê ~ *jVmcê* ‘nest’:
PCerr **jacê* > PNJ **jacê*; PNR *s-asê*; PCJ **ĵaci* (XAV *ĵaci*, XER *zasi*)
PSJ **jəŋθê* (vb. **jəŋθe-ñ*) > KGG *jaŋɸê* (vb. *jaŋɸe-ñ*)
PJ **-cĩ(C)* ‘bandage, sling’:
PCerr **-cĩ(C)* > PNJ **ja-cĩ, *ka-cĩ*
PSJ **-θĩ* (vb. *-θĩ-n*) ‘bundle, bandage’ > KGG *-ɸĩ* (vb. *-ɸĩ-n*); LKL *-ðẽ* (vb. *-ðẽ-n*)
PJ **kapnĵa* ‘to chew’:
PCerr **kapnĵa/*kapnĵa-r’* > PNJ **kapnĵa/*kapnĵa-r*; PCJ **waca/*waca-ri* (XER *wasa/wasa-r(i)*)
PSJ **kajã* > KGG *kajẽ*; LKL *ka[ŋ]jã/ka[ŋ]jã-ŋ*
PJ **nĵy* ‘bee, bumblebee’:
PCerr **(am-)nĵy* > PNJ **(am-)nĵy*; PCJ **am-ĵə* (XAV *am-ĵə* ‘*Tetragona clavipes*’, XER *am-zə* ‘Brazilian wasp (*Protonectarina sylveirae*)’)
PSJ **cy* > KGG *xy*; LKL *cy*
PJ **nĵô* ‘to hang.SG’:
PCerr **nĵô/*nĵôp-r’* > PNJ **nĵô/*nĵô-r*; PCerr **ja-nĵô/*ja-nĵôp-r’* ‘to hang.PL’ > PNJ **ja-nĵô/*ja-nĵô-r*; PCJ **ĵa-cõ/*ĵa-cõm-rĩ* (XAV *ĵa-cõ/ĵa-cõm-rĩ*, XER *za-sõ/za-sõm-r(ĩ)*)
PSJ **cə/*ca-ŋ* (vb. **cə-m*) > KGG *xa/xã-ŋ* (vb. *xa-m*);[33] LKL *co/ca-ŋ*
PJ **nĵê* ‘to tickle’:
PCerr **nĵê/*nĵê-r’* > PNJ **nĵê/*nĵê-r*
PSJ **ku-cê/*ku-ce-ŋ* > KGG *ku-xê/ku-xe-ŋ*; LKL *ku-cê/ku-ce-ŋ*
PJ **ja* ‘to stand.SG’:
PCerr **ja/*ja-m’* > PNJ **ĵa/*ĵã-m*; PNR *s-a:-ŋ*; PCJ **ĵa/*ĵam // *ĵa* (XAV *ĵa/ĵam // ĵa*, XER *za ~ da*)
PSJ **jã/*jã-ŋ* > KGG *jẽ/jẽ-ŋ*; LKL *jã* (vb. *jã-ŋ*)
PJ **ja* ‘to put vertically.SG’
PCerr **ja/*ja-r’* > PNJ **ĵa/*ĵa-r*; PNR *sa-rĭ*; PCJ **cə̃/*cə̃-rĩ* (XAV *cə̃/cə̃-rĩ*, XER *sə̃/sə̃-r(ĩ)*)
PSJ **jã* (vb. **jã-ŋ*) ‘to be put vertically’ > KGG *ɸ-ẽ* (vb. *ɸ-ẽ-ŋ*); LKL vb. *jã-ŋ*, 3 *ð-ã-ŋ*
PJ **ñĩja* ‘nose’:
PCerr **ñĩja-kre* > PNJ **ñĩja-kre*; (?) PCJ **ñĩci-krê* (XAV *ñĩci-krê*, XER *n-krê*)
PSJ **nĩjã* > KGG *nĩjẽ*; LKL *nẽjã*
PJ **jañ* ‘to eat.INTR, food’
PCerr **ĵaj’* > PCJ **caj // *ca* (XAV *caj // ca*, XER *saj // sa*)
PSJ **jã/*jã-ŋ/*jã-n* (vb. **jã-n*) > KGG **jẽ/*jẽ-ŋ/*jẽ-n* (vb. *jẽ-n*); LKL *jã-n*
PJ **jə̂* ‘bitter’:
PCerr **jə̂* > PNJ **ĵə̂*; PCJ **ĵe* (XAV *ĵe*, XER *ze*)
PSJ **θ-ə* (vb. **θ-a-ŋ*) > KGG *ɸ-a* (vb. *ɸ-õ-ŋ*); LKL *ð-o* (vb. *ð-a-ŋ*)
PJ **ñĩjə̂* ‘smoke’:
PCerr **ñĩjə̂* > PCJ **-ñĩĵe* (XAV *-ñĩĵe*, XER *-ñĩze*)
PSJ **nĩjə* (vb. **nĩja-ŋ*) > KGG *nĩja* (vb. *nĩjõ-ŋ*); LKL *nĩjo*
PJ **jʏ* ‘urine’:
PCerr **jə* > PNJ **ĵə*; PCJ **ĵe* (XAV *ĵe*, XER *ze* ‘bladder’)
PSJ **jə̂-ñ* ‘to urinate’ > KGG *jə-ñ*; LKL *jə-ñ*

[32] The expected form would be KGG **pe-ɸa* (vb. **pe-ɸõ-m*), given both the Laklãnõ cognate and the Kaingáng plural *pi-ŋ-ɸa* (vb. *pi-ŋ-ɸõ-m*).

[33] Stative KGG *xəw* ‘to hang’ is apparently a non-etymological formation, back-derived from *xa-m*.

PJ **ñỹ* ‘to sit.SG’:
PCerr **ñỹ/*ñỹp-r* > PNJ **ñỹ/*ñỹ-r*; PNR *sĩ-r/sĩ-rĭ*; PCJ **ñãm-rə̃* (XAV *ñãm-rə̃*, XER *nə̃m-rə̃*)
PSJ **nĩ* (vb. **nĩ-m*) > KGG *nĩ* (vb. *nĩ-m*); LKL *nẽ* (vb. *nẽ-m*)
PJ **jo* ‘tooth’:
PCerr **jwa* > PNJ **ĵwa*; PNR *swa*; PCJ **kwa* (XAV, XER *kwa*)
PSJ **ja* > KGG *jõ*; LKL *ja*
PJ **kVjo* ‘salt’:
PCerr **kVjwa* > PNJ **kaĵwa*; PCJ **kVkwa* ((?) XAV *ʔĩʔwa-wa:hə*, XER *kəkwa-rə*)
PSJ **kəja* (vb. **kəja-ŋ*) ‘salty, sour’ > KGG *kajõ* (vb. *kajõ-ŋ*)
PJ **jô* ‘to tear, to rip’:
PCerr **-jô/-jô-r’* > PNJ **ka-ĵô/*ka-ĵô-ñ*
PSJ **jə/*ja-ŋ, *jə-r* (vb. **ja-n*) > KGG *ja-r* (vb. *jõ-n*); LKL *ðã-jo/ðã-ja-ŋ, ðã-jol* (vb. *ðã-ja-n*)
PJ **jʊ* ‘pus’:
PCerr **jup-r* > PNJ **ĵur*; PCJ **ĵubruj // *ĵubru* (XAV *ĵubruj // ĵubru*, XER *zbruj // zbru*)
PSJ **θ-ô* (vb. **θ-o-m*) > KGG *ɸ-ô* (vb. *ɸ-o-m*); LKL *ð-ô* (vb. *ð-u-m*)
PJ **jʊ*, postposition:
PCerr **ju* > PNR *su:* ‘looking for’; PCJ **ĵô*[34] ‘looking for; for’ (XAV *ĵô*, XER *zô*)
PSJ **jô* ‘in front of’ > KGG *jô* ‘in front of; else’; LKL *jô* ‘in front of’
PJ **jʊm’* ‘father’:
PCerr **jum’* > PNJ **ĵũ(m)*; PNR *jum-pjə*
PSJ **joŋ* > KGG *joŋ*; LKL *juŋ*
PJ **peju* ‘to hide’:
PCerr **pju/*pju-r’* > PNJ **puĵu/*puĵu-r*
PSJ **peju* > KGG *peju*; LKL *peju*
PJ **nĵu* ‘to heat, to dry’:
PCerr **nĵu/*nĵu-ñ’* > PNJ **nĵu/*nĵu-ñ* ‘to dry in the sun’; Southern Kayapó ⟨timuçúnkuátú, kuatáunçún⟩ (likely *nsu-ŋ*)
PSJ **cu-r* (vb. **cu-n*) ‘to heat by the fire’ > KGG *cur* (vb. *cu-n*); LKL *cul* ‘cooked half’; PSJ **jaŋ-cu/*jaŋ-cu-ŋ, *jaŋ-cu-r* (vb. **jaŋ-cu-n*) ‘to roast’ > KGG *jõŋ-xu/jõŋ-xu-ŋ, jõŋ-xu-r* (vb. *jõŋ-xu-n*); LKL *jaŋ-cu/jaŋ-cu-ŋ, jaŋ-cu-l* (vb. *jaŋ-cu-n*)
PJ **ũt/*ñũt* ‘to sleep’:
PCerr **ŋõt/*ñõt* > PNJ **ŋõr/*ñõt*; PNR *s-õtĭ*; PCJ **ñõtõ // *ñõ:nõ* (XAV *ñõtõ // ñõ:nõ*, XER *nõt(õ) ~ ntõ*)
PSJ **nũr* > KGG *nũr*; LKL *nũl*
PJ **jê (~ *jên)* ‘thread, knot, cloth’:
PCerr **jê (~ *jên)* > PNJ **ĵê*; PNR *sê*
PSJ vb. **θe-n* ‘to spin thread’ > KGG *ɸe-n*; LKL *ðe-n*
PJ **ñũmjê(C)* ‘female breast’:
PCerr **ñõmĵê(C)* > PNJ **ñõmĵê*
PSJ **nũŋjê* ‘female breast, milk’ > KGG *nũŋjê*; LKL *nũŋnĵê*
PJ **kujêk* ‘vein’:
PCerr **kujêk* > PNJ **kuĵêk*
PSJ **kujêj* > KGG, LKL *kujêj*
PJ **ji* ‘to lay.SG’:
PCerr **ji/*ji-r’* > PNJ **ĵi/*ĵi-r*; PCJ **hi/*hi-ri* (XAV *hi/hi-ri*, XER *hi/h(i)-ri*)
PSJ **ji/*ji-ŋ* > KGG *ɸ-i/ɸ-i-ŋ*; LKL *ji/ji-ŋ*

[34] The expected reflex would be PCJ **ĵu*. Note that the vowel *ô* is extremely rare in Central Jê, its occurrence being restricted to a few function words.

PJ **jiji ~ *ñĩji* ‘name’:
PCerr **ñĩji* > PNJ **ñĩ-ĵi*; PNR *issi*; PCJ **ñĩci // *ñĩːci* (XAV *ñĩci // ñĩːci*, XER *nĩsi-ze*)
PSJ **jiji ~ *jyjy* (vb. **jiji-n ~ *jyjy-n*) > KGG *jiji ~ jyji ~ jyjy* (vb. *jiji-n*); LKL *jyjy*, 3 *ð-yjy* (vb. *jyjy-n*, 3 *ð-yjy-n*)
PJ **ñĩ* ‘meat’:
PCerr **ñĩ* > PNJ **ñĩ*; PNR *ñĩ*; PCJ **ñĩ* (XAV *ñĩ*, XER *nĩ*)
PSJ **nĩ* > KGG *nĩ*; LKL *nẽ*
PJ **ñĩ(m)-* ‘hand (*in compounds*)’:
PCerr **ñĩ(m)-kra* ‘hand’ > PNJ **ñĩ-kra ~ *ñỹ-kra*; PNR *s-i-kja*; PCJ **ñĩp-kra* (XAV *ñĩp-ʔrata // ñĩp-ʔraːda*, XER *nĩp-kra*)
PSJ **nĩ-* > KGG *nĩ-* (as in *nĩ-pê* ‘to wash hands’, *nĩ-ju/nĩ-ju-ŋ* ‘to show with one’s hand’, *nĩ-ŋge* (vb. *nĩ-ŋõ-n*) ‘hand’); LKL (as in *nẽ-ju-jo* ‘index finger’, *nẽ-ŋga* ‘hand’)
PJ **kaj*’ ‘basket’:
PCerr **kac* > PNJ **kac*; (?) PNR *kaŋ*
PSJ **kãj* (vb. **kãñ*) > KGG *kẽj*; LKL *kãj*
PJ **kyñ* ‘skin, bark’, **jar-kyñ* ‘lip’:
PCerr **kyj*’ ‘skin, bark’ > PNJ **kə*; PNR *kə*; PCJ **həj // *hə* (XAV *hə*, XER *həj ~ hê // hə*); PCerr **jaj-kyj*’ ‘lip’ > PNJ **ja-kə*; PNR *s-a-kə*; PCJ **ĵaj-həj // *ĵaj-hə* (XAV *ĵaj-həj // ĵaj-hə*, XER *zaj-hə* ‘buttocks’)
PSJ **jãn-ky* (vb. **jãn-ky-n*) ‘mouth’ > KGG *jẽn-ky* (vb. *jẽn-ky-n*); LKL *jãn-ky*
PJ **ko* ‘to dig’:
PCerr **kwa/*kwa-ñ*’ > PNJ **kwə̂/*kwə̂-ñ*; PNR *kwə-ŋ*
PSJ **ka/*ka-ŋ* (vb. **ka-n*) > LKL *ka/ka-ŋ* (vb. *ka-n*)
PJ **jar-ko* ‘mouth’:
PCerr **jad-kwa* > PNJ **jar-kwa*; PNR *s-a-kua*; PCJ **ĵada-wa* (XAV *ĵada-wa*, XER *zda-wa* ‘mouth, door’)
PSJ **jãn-ka* ‘door’ > KGG *jõn-kõ*; LKL *ð-ãn-ka*
PJ **kVñko* ‘sky’:
PCerr **kəjkwa* > PNJ **kəjkwa*; PCJ **həjwa* (XAV *həjwa*, XER *həjwa ~ hewa*)
PSJ **kəñka* > KGG *kañkõ*; LKL *koñka*
PJ **kôp* ‘fly, mosquito’:
PCerr **kôp* ‘fly’ > PNJ **kôp*; PCJ **kupu // *kuːbu* (XAV *ʔupu // ʔuːbu*, XER *kpu // kbu*)
PSJ **kə* ‘mosquito’ > KGG *ka*; LKL *cẽ-to-ko* ‘black fly’
PJ **kôm* ‘tree, horn’; **ñĩ-kôm* ‘horn’:
PCerr **kôm* > PNJ **kô*; PCJ **kõmõ // *ku* ‘horn’ (XAV *ʔõmõ // ʔu*, XER *kmõ // ku*); PCerr **ñĩ-kôm* ‘horn’ > PNJ **ñĩ-kô*
PSJ **kə* ‘tree’ > KGG *ka*; LKL *ko*; PSJ **nĩ-kə* ‘horn’ > KGG *nĩ-ka*; LKL *nẽ-ko*
PJ **kôk ~ *ka-kôk* ‘wind’:
PCerr **kôk, *ka-kôk* > PNJ **kôk*; PCJ **wa-kuku // *wa-ku* (XAV *-wa-ʔuʔu // -wa-ʔu*, XER *wa-kku*)
PSJ **kã-kə* (vb. **kã-kə-n*) > KGG *kõ-ka* (vb. *kõ-ka-n*)
PJ **kʊ* ‘to eat’:
PCerr **ku/*ku-r*’ > PNJ **ku/ku-r* ‘to eat.PL’; PNR *ku-rĭ*; PCJ **hu/*hu-ri* (XAV *hu/hu-ri*)
PSJ **kô* ‘to eat, to use’ > KGG, LKL *kô*
PJ **kẽt* ‘stone’:
PCerr **kẽt* > PNJ **kẽn*; PNR *kĩẽj*; PCJ **kẽtẽ // *kẽːnẽ* (XAV *ʔẽtẽ // ʔẽːnẽ*, XER *ktẽ // knẽ*)
Ingain *kẽr*: *Lista* ⟨queré, quené⟩, *María Antonia* ⟨keré⟩, *Cosme Román* ⟨kiné⟩, *Vogt*[3] ⟨kiré⟩
PJ **kêt* ‘NEG’:
PCerr **kêt* > PNJ **kêt*
PSJ **kêr* ‘ADVT’ > KGG *kêr*; LKL *kêl-o*

PJ **kɪj* ‘pit’
PCerr **kjaj’* > PNJ **kî* ‘earth oven’; PNR *kjê* ‘earth oven’
PSJ **kêj* ‘tomb’ > KGG *kêj*
PJ **kɪ ~ *kɪj ~ *kɪj’* ‘to split’
PCerr **-kja/*-kja-r’* > PNJ **ko-kî/*ko-kjê-r*
PSJ vb. **ke-ñ* > KGG *ke-ñ*; PSJ vb. **ka-ke-ñ* ‘to split; canoe’ > KGG *kõ-ke-ñ*; LKL *ka-ke-ñ*
PJ **kra* ‘offspring’:
PCerr **kra* > PNJ **kra*; PCJ **kra: // *kra* (XAV *ʔra: // ʔra*, XER *kra*)
PSJ **krã* (vb. **krã-ŋ*) > KGG *krẽ* (vb. *krẽ-ŋ*); LKL *krã*
PJ **krat* ‘macaw’:
PCJ **krata // *kra:da* (XAV *ʔrata // ʔra:da*, XER *krda*)
Ingain *kla(r)*: *Pedro* ⟨kluá⟩, *María Antonia* ⟨klán⟩, (?) *Vogt*[1] ⟨kakladeĩn⟩, *Vogt*[3] ⟨klá⟩
PJ **krỹñ* ‘head’:
PCerr **krãj’* > PNJ **krõ*; PNR *kjõ*; PCJ **krãj // *krã* (XAV *ʔrãj // ʔrã*, XER *krãj ~ krẽ // krã*)
PSJ **krĩ* (vb. **krĩ-n*) > KGG *krĩ* (vb. *krĩ-n*); LKL *klẽ*
PJ **jVkrỹñ* ‘knee’:
PCerr **jikrãj’* > PCJ **hikrãj // *hikrã* (XAV *hiʔrã-ti*, XER *hi-krãj-ti ~ hi-krẽ-ti, hi-krãj-*)
PSJ **jəkrĩ* > KGG *jakrĩ*; LKL *joklẽ*
PJ **krot* ‘chin, beak’:
PCerr **krwat* ‘beak’ > PNJ **krwât*; PCJ **wata // *wa:da* (XAV *wata // wa:da* ‘chin, beak’, XER *wda*)
Ingain *jat-krar* ‘chin’: *Lista* ⟨amincrará⟩, *María Antonia* ⟨miet krará⟩, *Vogt*[3] ⟨jitkyrará⟩
PJ **kre* ‘hole’:
PCerr **kre* > PNJ **kre*; PNR *kre*; PCJ **krê* ‘vagina, anus’ (XAV *ʔrê*, XER *krê*), **am-krê* ‘hole’ (XAV *am-ʔrê*, XER *am-krê*)
PSJ **kre* > KGG *kre*; LKL *kle*
PJ **kre* ‘to plant’:
PCerr **kre* > PNJ **kre*; PNR *kre*; PCJ **krê* (XAV *ʔrê*, XER *krê*)
PSJ **kre* (vb. **kra-n*) > KGG *kre* (vb. *krõ-n*); LKL *klel*[35] (vb. *kla-n*); PSJ **ã-kre/*jə-kre* ‘plantation’ (vb. **ã-kra-n/*jə-kra-n*) > KGG *ẽ-kre/ja-krõ-n*
PJ **krɪñ* ‘thigh’:
PCerr **krjaj’* > PNJ **kjê*; PCJ **ĵaj // *ĵa* (XAV *ĵaj // ĵa*, XER *zda*)
PSJ **krê* > KGG *krê*; LKL *klê*
PJ **ŋə̃(C)* ‘to push against, to grind’:
PCerr **ŋə̃/*ŋə̃-ñ’* > PNJ **ŋə̃/*ŋə̃-ñ*
PSJ **ŋə̃/*ŋə̃-ŋ* > KGG *ŋə̃*
PJ **ŋgoñ’* ‘feather, hair’:
PCerr **ŋgwañ’* > PNJ **ŋgwôñ* ‘feather’; PNR *iŋkwəŋ*
PSJ **ŋgañ* ‘animal hair’ > KGG *ŋõñ* ‘hair’; LKL *ŋgañ* ‘horsehair’
PJ **ŋgô* ‘louse’:
PCerr **ŋgô* > PNJ **ŋgô*; PNR *kjə̃-ŋkô*; PCJ **ku* (XAV *ʔu*)
PSJ **ŋgə* (vb. **ŋga-n*) ‘louse, flea’ > KGG *ŋga* (vb. *ŋõ-n*); LKL *ŋgo* (vb. *-ŋga-n*)
PJ **ŋũ* ‘to push against, to crumble’:
PCerr **-ŋõ/*-ŋõ-ñ’* > PNJ **ka-ŋõ/ka-ŋõ-ñ*
PSJ **ŋũ/*ŋũ-ŋ* > KGG *ŋũ/ŋũ-ŋ*

[35] Final *-l* in the Laklãnõ form is apparently non-etymological. It might have arisen as a back-derivation from the regular verbalized form *kla-n*.

PJ *ŋgɪ ‘to enter.PL’:
PCerr *a-ŋgja/*ŋgja-c > PNJ *a-ŋgî/*ŋgjê-c; PNR iŋkjə; PCJ *ə̃:-ĵa/ĵa-ci ‘to enter.DU’ (XAV ʔə̃:-ĵa/ĵa-ci, XER za-s(i))
PSJ *ŋgê (vb. *ŋgê-m) > KGG ŋgê (vb. ŋgê-m); LKL ŋgê (vb. ŋgê)
PJ *ŋrũ (~ *ŋrũn) ‘toucan’:
PCerr *ŋrõ > PNJ *ŋrõ; PNR iŋkjõ-pepetĭ
PSJ *ŋrũ > KGG ŋrũ; LKL ŋlũ
PJ *ŋgre ‘egg’:
PCerr *ŋgre > PNJ *ŋgre; PNR iŋkre; PCJ *krê (XAV ʔrê, XER krê)
PSJ *ŋgre ‘egg, penis’ > KGG ŋgre ‘penis’; LKL ŋgle
PJ *ŋgre ‘to dance’:
PCerr *ŋgre/*ŋgre-r > PNJ *ŋgre/*ŋgre-r; PNR iŋkre:; PCJ *aj-krẽ/*ci-krẽ-nẽ (XAV aj-ʔrẽ/aj-ʔrẽ-nẽ ~ ci-ʔrẽ-nẽ)
PSJ *ŋgre (vb. *ŋgre-n) > KGG vb. ŋgre-n; LKL ŋgle (vb. ŋgle-n)
PJ *ŋrĩ(C) ‘wrap’:
PCerr *-ŋrĩ(C) > PNJ *ku-ŋrĩ ‘to gather in a bundle’, *-ŋrĩ ‘to make packages’
PSJ *ŋrĩ/*ŋrĩ-ŋ, *ŋrĩ-r (vb. *ŋrĩ-n) > KGG ŋrĩ/ŋrĩ-ŋ/ŋrĩ-n; LKL ŋlẽ/ŋlẽ-ŋ, ŋlẽ-r (vb. ŋlẽ-n)

Semantic issues

PJ *pu:
PCerr *pu > PNJ *pu ‘tube (in compounds)’; PCerr *ŋgrwa-pu ‘moriche stem’ > PNJ *ŋgrwa-pu, PCJ *wa:-bu (XAV wa:-bu, XER wa-bu)
PSJ *pu (vb. *pu-ŋ) ‘handle’ > KGG pu (vb. pu-ŋ); LKL vb. pu-ŋ
PJ *prĩt:
PCerr *prĩt > PNJ *prĩt ‘pequi’
PSJ *prĩr > KGG prĩr ‘araucaria resin’; (?) LKL plẽl ‘reddish’
PJ *wẽ(C):
PCerr *wẽ(C) > PNJ *bẽ/*bẽ-r ‘to show’, *bẽ-r ‘speech’; PNR pẽj/pẽj-ñ ‘to say’
PSJ *wĩ (vb. *wĩ-n) ‘to speak’ > KGG wĩ (vb. wĩ-n); LKL vẽ
PJ *mẽ:
PCerr *mẽ/*mẽ-ñ’ ‘to throw.SG’ > PNJ *mẽ/*mẽ-ñ; PNR mẽ-ñ/mẽ:-rĭ; PCJ *mẽ/*mẽ-j // *mẽ (XAV mẽ/mẽ-j // mẽ, XER mẽ)
PSJ *mĩ/*mĩ-ŋ/*mĩ-n ‘spill’ > KGG mĩ/mĩ-ŋ/mĩ-n
PJ *tɪt:
PCerr *tjat ‘to burn’ > PNJ *tjêr/*tjêt; PNR titĭ; PCJ *ĵata // *ĵa:da (XAV ĵata // ĵa:da, XER zata)
PSJ *têr (vb. *te-n) ‘to die.ACT’ > KGG têr (vb. te-n); LKL têl (vb. te-n)
PJ *rət:
PCerr *-rot > PNJ *ñĩ-rot ‘crooked’
PSJ *rãr ‘sharp-ended; thorn’ > KGG rẽr; LKL lãl
PJ *ro:
PCerr *rwa/*rwa-k > PNJ *rû/*rwô-k ‘to descend’
PSJ *ra ‘to enter.SG, to begin.SG’ > KGG rõ; LKL la
PJ *kʊrôt:
PCerr *kurôt > PNJ *kurôt ‘skin irritation’ (only Apinajé)
PSJ *kôrər > KGG kôrar ‘frog’
PJ *rê:
PCerr *rê/*rê(p)-r > PNJ *rê/*rê-r ‘to cross’
PSJ *rê (vb. *rê-m) ‘to descend.PL’ > KGG *rê (vb. rê-m); LKL lê (vb. lê-m)

PJ **nĵa* ‘to bite’:[36]
PCerr **nĵa/*nĵa-r* ‘to bite’ > PNJ **nĵa/*nĵa-r*; PNR *insa-rĭ*; PCJ **ca/*ca-ri* (XAV *ca/ca-ri*, XER *ca/ca-r(i)*)
PSJ **cã* ‘game *(animals)*; war’ > KGG *xẽ*
PJ **nĵʊm’*:
PCerr **nĵum’* > PNJ **nĵũm* ‘dirty’
PSJ **coŋ* ‘egg yolk; larvae’s food’ > KGG *xoŋ*; (?) LKL *cuŋ* ‘bee sp. (*Plebeia sp.*)’
PJ **nĵi(C)*:
PCerr **nĵi(C)* > PNJ **nĵi* ‘mother’; (?) PNR *si-pjə* ‘wife’
PSJ **ci* ‘old’ > KGG *xi*; LKL *ci*
PJ **ŋgrê*:[37]
PCerr **ŋgrê* > PNJ **ŋgrê* ‘few’
PSJ **ŋgrê/*ŋgre-ŋ* ‘to sieve’ > KGG *ŋgrê/ŋgre-ŋ*; LKL *ŋglê/ŋgle-ŋ*

Phonetic issues

PJ **pu ~ *pʊ* (if PSJ is exceptional) ~ **po* (if PCerr is exceptional) ‘to wrap’:
PCerr **ku-pu* > PNJ **ku-pu*; PCJ **ku-bu* (XAV *ʔu-bu*)
PSJ **pa* (vb. **pa-ŋ*) > KGG *põ* (vb. *põ-ŋ*); LKL *pa* (vb. *pa-ŋ*)
PJ **wʏr’* (if PSJ is exceptional) ~ **wôr’* (if PNJ/PCerr is exceptional) ‘tree sp.’:
PCerr **wəd ~ *wôd* ‘tree’ > PNJ **bər*; PNR *pərĭ*; PCJ **wêdê // *wê:dê* (XAV *wêdê // wê:dê*, XER *wdê*)
PSJ **wan* ‘bamboo’ > KGG *wõn*; LKL *van*
PJ **tʏm’* (if PSJ is exceptional) ~ **tôm’* (if PNJ/PCerr is exceptional) ‘new’:
PCerr **təm’ ~ *tôm’* > PNJ **təm* ‘raw’; PCJ **tem // *te* (XAV, XER *tem // te*)
PSJ **taŋ* > KGG *tõŋ*; LKL *taŋ*
PJ **tʊt* (if PSJ is exceptional) ~ **tʊn’* (if PNJ/PCerr is exceptional) ‘pigeon’:
PCerr **tut* > PNJ **tut*
PSJ **tôn* > KGG *tôn*
PJ **jar-râ* (if PSJ is exceptional) ~ **jar-rʏ* (if PNJ/PCerr is exceptional) ‘saliva’:
PCerr **jad-râ ~ *jad-rə* > PNJ **jar-rə*; PCJ **ĵare* (XER *zdare mẽ* ‘to spit’ = ‘to throw saliva’)
PSJ **jarə* (vb. **jara-n*) > KGG *jõra* (vb. *jõrõ-n*); LKL *jãlo*
PJ **jVmṼ* ‘cheek’:
PCerr **jamṼ* > PNJ **jamã*
PSJ **jəmã* > KGG *jamẽ*; LKL *jomã*
PJ **kâ* (if PSJ is exceptional) ~ **kʏ* (if PNJ/PCerr is exceptional) ‘bad smell’:
PCerr **kâ* > PNJ **kâ*
PSJ **kʏ* > KGG, LKL *kʏ*
PJ **ndon* (if PSJ is exceptional) ~ **ndun* (if PNJ/PCerr is exceptional) ‘snail’:
PCerr **ndwan’* > PNJ **ndwôn*; PNR *intôw*
PSJ **ndun* > KGG *ndôn nũnẽ, ndun*; LKL *ndun* ‘caterpillar’
PJ **kʊm’* (if PSJ is exceptional) ~ **ŋgʊm’* (if PNJ/PCerr is exceptional) ‘mist, cloud’:
PCerr **kum’ ~ *ŋgum’* > PNJ **kũm* ‘smoke, mist’; PCJ **hum // *hu*; PCerr **ka-kum’* ‘cloud, dry season’ > PNJ **ka-kũm* ‘cloud’; (?) PNR *a-kuŋ* ‘dry season’; (?) PCJ **wa-hum // *wa-hu* ‘dry season, year’

[36] Despite the semantic divergence between the Cerrado languages and Kaingáng, which makes the comparison dubious, the Proto-Cerrado root may safely be identified as a retention from Proto-Jê, given that a secure external cognate exists in Maxakalí: /ca/ ‘to bite, to sting’.

[37] If this comparison is correct, the Proto-Jê word could have meant ‘small’ and correspond externally to Maxakalí /ktĩk-nãk/ ‘small’.

PSJ **ŋgoŋ* ‘cloud’ > KGG *ŋgoŋ*; LKL *ŋguŋ*

PJ **ŋgôj* (if PSJ is exceptional) ~ **ŋgʊj* (if PNJ/PCerr is exceptional) ‘water’

PCerr **ŋgôj ~ *ŋguj* > PNJ **ŋgô*; PNR *iŋkô*; PCJ **kuj // *ku* ‘still water’ (XAV *ʔuy // ʔu*, XER *kuj-*)

PSJ **ŋgôj* (vb. *ŋgoñ*) > KGG *ŋgôj* (vb. *ŋgoñ*); LKL *ŋgôj*

Reflexes only in PCerr

PCerr **pôr/*pôk* ‘to ignite’ > PNJ **pôr/*pôk*; PNR *pô* ‘to burn’

Possible PJ reconstruction: **pôk*. External cognate: Maxakalí /pyk/ ‘to burn (vi)’, Rikbáktsa *pok*, Proto-Tupian **puk*.

PCerr **a-pê/*jV-pê-ñ*’ ‘to make’ > PNJ **a-pê/*jə-pê-ñ* ‘to work’; PNR *s-ə-pê* ‘to work’; PCJ **ap-pi/*ñĩp-pi* (XAV *ap-pi/ñĩp-pi* ‘to cook’, XER *ni-pi*)

Possible PJ reconstruction: **-pê*. External cognate: Krenák *pi*.

PCerr **kupê/*kupê-ñ*’ ‘to touch’ > PNJ **kupê/*kupê-ñ*; PCJ **kupi* (XAV *ʔupi*, XER *kupi*)

Possible PJ reconstruction: **-pê*. External cognate: Maxakalí /REL-pit/.

PCerr **prâ(j’)* ‘feather, corn husk’ > PNJ **prâ*

Possible PJ reconstructions: **prâ, *prâj,* or **prâñ*. External cognate: Maxakalí /ptyc-nãk/ ‘bird’.

PCerr **mbra*[38] ‘to walk.PL’ > PNJ **mbra/*mbra-r*; PCJ **kmə̃-pra/*kmə̃-pra-ba* ‘to take away’ (XAV *ʔmə̃-pra/ʔmə̃-pra-ba*, XER *kmə̃-pra/kmə̃-pra-ba ~ kmə̃-pra-b ~ kmə̃-pr-ba*), **pra-ba* ‘to run’ (XAV *pra-ba*, XER *pra-ba ~ pra-b ~ prb-a* ‘to dance’)

Possible PJ reconstruction: **mbra(C)*. External cognate: Rikbáktsa *parak* ‘to walk, to run.SG’.

PCerr **mbwa/*mbwa-r*’ ‘to cry’ > PNJ **mbû/*mbâ-r*

Possible PJ reconstruction: **mbo*. External cognates: Maxakalí /pu-t-a, IRR pu/ (Ritual Maxakalí /pu-t, IRR pu/), Krenák *pu-k*, Karajá *bu*, Proto-Jabutí **mbo*.

PCerr **-mrə̃(m)* ‘to bathe (vt.)’ > PNJ **kumrə̃*

Possible PJ reconstruction: **-mrỹ* or **-mrỹm*. External cognates: Maxakalí /mnỹp/ ‘to submerge’, (?) Rikbáktsa *para* (unless related to PSJ **mbrô* ‘to swim, to bathe’ > KGG *mbrô*, LKL *mblô*).

PCerr **mrõ(C)* ‘to dive, to soak; to cook’ > PNJ **mrõ*

Possible PJ reconstruction: **mrũ(C)* or **mrõ(C)*. External cognates: Rikbáktsa *mũrũ* ‘to bathe (vi)’, Proto-Jabutí **mrə̃* ‘to bathe (vi)’.

PCerr **wə̃/*wə̃(p)-r* ‘to smell’ > PNJ **bə̃/*bə̃-r*; (?) Southern Kayapó ⟨tipén⟩ (likely *pə̃*)

Possible PJ reconstruction: **wə̃*. External cognates: Maxakalí /cy-pyp/ ‘nose’, /ca-pyp/ ‘pig’, Ritual Maxakalí /byp/ ‘to smell’, /ca-byp/ ‘pig’, Krenák *wəp* ‘to smell, to kiss’.

PCerr **wôc* ‘to arrive’ > PNJ **bôj/*bôc*; PNR *pôw/pôw-rĭ*; (?) PCJ **wi/*wi-ci* ‘to arrive.SG’ (XAV *wi/wi-ci*, XER *wi/w(i)-si*)

Possible PJ reconstruction: **wôc*. External cognate: Proto-Tupian **wuc* ‘to arrive (Jurúna), to go out (Káro, Puruborá)’.

PCerr **wê* ‘ABL’ > PNJ **bê* ‘ABL, MALEF, COP’; PNR *pê:*; PCJ **wi* (XAV, XER *wi*)

Possible PJ reconstruction: **wê*. External cognate: Proto-Tupian **wi*.

PCerr **jawê* ‘to love, to respect’ > PNJ **jabê* ‘to love, to be melancholy’; PCJ **ĵawi* (XAV *ĵawi*, XER *zawi*)

Possible PJ reconstruction: **jawê*. External cognate: Maxakalí /capit/ ‘to call, to invite’.

PCerr **wĩ/*wĩ-r* ‘to kill.SG’ > PNJ **bĩ/*bĩ-r*; PNR *pĩ-rĭ*; PCJ **wĩ/*wĩ-rĩ* (XAV *wĩ/wĩ-rĩ*, XER *wĩ/w(ĩ)-rĩ*)

Possible PJ reconstruction: **wĩ*. External cognates: Maxakalí /mĩ-k, IRR mĩ/ ‘to make’ (Ritual Maxakalí ‘to kill’), Proto-Tupian **wĩ* ‘to kill’.

[38] It is unclear how the non-finite form of this verb should be reconstructed.

PCerr **tõj** 'brother' > PNJ **tõ*; PNR *tõ*; PCJ **nõj // *nõ* 'younger sibling of the same sex' (XAV *nõj // nõ*)

Possible PJ reconstruction: **tũj, *tõj, *tũñ,* or **tõñ*. External cognate: Maxakalí /nũc/.

PCerr **te* 'GEN, ERG' > PNJ **te* 'ERG, GEN (material, stimulus)'; PCJ **tê* (XAV *tê* '1SG.ERG', *wa-tê* '1PL.ERG', *(tê-)tê* '3.ERG', XER *tê*)

Possible PJ reconstruction: **te*. External cognates: Maxakalí /te/, (?) Krenák *ti* 'I', *ho-ti* 'you'.

PCerr **tik* 'belly' > PNJ **tik*; PCJ **diki // *di* (XAV *diʔi // di*, XER *dki // di*)

Possible PJ reconstruction: **tik*. External cognates: Maxakalí /tek/.

PCerr **ndap* 'sour' > PNJ **ndap* 'sour, ripe'; PCJ **wam-tapa* 'sour, bitter' (XER *wam-tap(a) ~ wam-tpa*)

Possible PJ reconstruction: **ndap*. External cognate: Proto-Tupian **ndəp*.

PCerr **ndep* 'ripe' > PNJ **ndep*

Possible PJ reconstruction: **ndep*. External cognate: Tuparí *tep* (if from Proto-Tupian **ndep*).

PCerr **pʾrək* 'to look like' > PNJ **pyrək*

Possible PJ reconstruction: **pVrʏk*. External cognates: Maxakalí /pytyk/.

PCerr **rã(r)* 'flower' > PNJ **rã*; PNR *ijã*; PCJ **ñĩ-rãrã ~ *ñĩ-rãnã* (XAV *ñĩ-rãrã*, XER *nĩ-rãn(ã) ~ nĩ-rnã*)

Possible PJ reconstruction: **rã* or **rãr*. External cognate: Maxakalí /-dyt/.

PCerr **kucym* 'fire' > PNJ **kucy*; PNR *issy*; PCJ **kuñãmã // *kuĵə* (XAV *ʔuñãmã // ʔuĵə*, XER *kunmã // kuzə*)

Possible PJ reconstruction: **kucym*. External cognates: Maxakalí /kycap/, Karajá *he-koɗy*.

PCerr **ci* 'bone' > PNJ **ci*; PCJ **hi* (XAV, XER *hi*)

Possible PJ reconstruction: **ji* or **ci*. External cognates: Krenák *jek*, Karajá *di*, Ofayé *hi ~ hih*, Rikbáktsa *ek* 'leg', Jabutí **ĵi, *i*, Chiquitano *-i* 'leg, plant', *-pa-ʔi* 'bone'.

PCerr **jəñ'* 'sweet, tasty' > PNJ **ĵəñ'*; PCJ **ĵêj // *ĵê* (XAV *ĵêj // ĵê*, XER *zêj // zê*)

Possible PJ reconstruction: **jʏñ'*. External cognates: Maxakalí /cyc-pek/, Tuparí *hoc* (if from Proto-Tupian **joc ~ *jôc*).

PCerr **jwañ'* 'NMLZ.AG' > PNJ **ĵwâñ*, PCJ **-kwaj // *-kwa*

Possible PJ reconstruction: **joñ'*. External cognate: Karajá /-du/.

PCerr **jajwa/*jajwa-r'* 'to lay.PL, to spill' > PNJ **jaĵû/*jaĵwâ-r*; PNR *s-aswə-rĭ* 'to spill'; PCJ **c-a(j)kwa/*c-a(j)kwa-ri* (XAV *c-aʔwa/c-aʔwa-ri*, XER *s-ajkwa/s-ajkwa-r(i)*)

Possible PJ reconstruction: **-jo*. External cognate: Maxakalí /nỹ=cu-k, IRR nỹ=cu/.

PCerr **jwa(m)/*jwa(m)-r* 'to bathe' > PNJ **ĵû/*ĵwâ-r*; PNR *swə-rĭ*

Possible PJ reconstruction: **jom*. External cognate: Krenák *jum*, Proto-Jabutí **dzo*.

PCerr **pʾji* 'one' > PNJ **pyĵi*; PCJ **pici* (XER *pici* 'only')

Possible PJ reconstruction: **pVji*. External cognates: Maxakalí /pycet/, (?) Krenák *pucik*.

PCerr **ñṼt* 'faeces, guts' > PNJ **ñĩn*; PNR *ñĩ:*; PCJ **ñãnã // *ñã:nã*

Possible PJ reconstruction: **ñṼt*. External cognates: Maxakalí /ñũt/, Chiquitano *-ãʔã*.

PCerr **kwar* 'manioc' > PNJ **kwâr*; PNR *kwy*

Possible PJ reconstruction: **kor*. External cognate: Maxakalí /kut/.

PCerr **kwa(C)/*kwa(C)-r* 'to defecate' > PNJ **ij-kû/*kwâ-r*

Possible PJ reconstruction: **ko(C)*. External cognates: Karajá *ku*.

PCerr **kryt* 'metal, flint' > PNJ **kryt*; PCJ **hətə // *hɔ:də* (XAV *hədə*, *hətə-ra* 'ax', XER *hdə*)

Possible PJ reconstruction: **kryt*. External cognate: Krenák *krak* 'metal, knife'.

PCerr **kroj'* 'rotten' > PNJ **kro*; PCJ **kroj // *kro* (XAV *ʔroj // ʔro*, XER *kroj // kro*)

Possible PJ reconstruction: **krəj* or **krəñ*. External cognates: Maxakalí /ktuc/, Rikbáktsa *horo*.

PCerr **krẽ/*krẽ-r* 'to eat.SG' > PNJ **krẽ/*krẽ-r*; PCJ **krẽ/*krẽ-nẽ* (XAV *ʔrẽ/ʔrẽ-nẽ*, XER *krẽ/kr(ẽ)-nẽ*)

Possible PJ reconstructions: **krẽ* or **krẽŋ'*. External cognate: Maxakalí /knỹk/ 'to have sex with'.

PCerr *ŋgryk ‘angry’ > PNJ *ŋgryk; PCJ *həkə // *hə (XAV həʔə // hə)
Possible PJ reconstruction: *ŋgryk. External cognate: Krenák *ŋgraŋ* ‘angry, snake’.
PCerr *uɣa/*jəp-r ‘to roast’ > PNJ *ga/*ĵə-r; PCJ *ĵêb-rê (XAV *ĵêb-rê*, XER *dêb-r(ê)*)
Possible PJ reconstruction: *a (*/ap’/). External cognates: Maxakalí /mũ=hap/ ‘to roast, to burn’, Krenák *op*.
PCerr *ŋõ/*ñõp-r’ ‘to give’ > PNJ *ŋõ/*ñõ-r; PNR *s-õ-rĭ*; PCJ *cõ/*cõm-rĩ (XAV *cõ/cõm-rĩ*, XER *sõ/sõm-r(ĩ)*)
Possible PJ reconstruction: *ũ or *õ (*/ũp’ ~ õp’/). External cognates: Maxakalí /hũp/, Krenák *um̥*, Karajá *õ*, Proto-Jabutí *ũ*.

Reflexes only in PSJ

PSJ *-pê/*-pe-ŋ ‘to wash’ > KGG *-pê/-pe-ŋ*; LKL *-pê/-pe-ŋ*
Possible PJ reconstructions: *-pê or *-pɪ. External cognates: Maxakalí /pi-k, IRR pi/, Rikbáktsa *pik*, Proto-Jabutí *pi*.
PSJ *prə/*pra-ŋ ‘to bite’ > KGG *pra/prõ-ŋ*; LKL *plo/pla-ŋ*
Possible PJ reconstructions: *prô, *prôp, *prôm, *prâ, *prâp, or *prâm. External cognates: Maxakalí /ptup/, Karajá *ro*, Rikbáktsa *boro*.
PSJ *mãŋ ‘bee, honey’ > KGG *mãŋ*; LKL *mõŋ*
Possible PJ reconstruction: *mãŋ’. External cognates: Maxakalí /pyk/, Krenák *pəŋ*, Ofayé *ɸok ~ ɸak ~ ɸək ~ ɸəgⁿ ~ ɸok-təj ~ ɸəg-təj*.
PSJ *mbeŋ ‘ax’ > KGG, LKL *mbeŋ*
Possible PJ reconstructions: *mbêm’ or *mbɪm’. External cognate: Maxakalí /pip-/.
PSJ *wə̂ ‘rotten’ > KGG *wə*
Possible PJ reconstruction: *wʏm. External cognates: Krenák *wəm*, (?) Ofayé *ɸa:ʔ*.
PSJ *reŋ-rê (vb. *reŋ-re-ŋ) ‘two’ > KGG *reŋ-rê* (vb. *reŋ-re-ŋ*); LKL *leŋ-lê* (vb. *leŋ-le-ŋ*)
Possible PJ reconstructions: *-rê or *-rɪ. External cognate: Maxakalí /tik/.
PSJ *rir ‘soft (of corn)’ > KGG *rir*
Possible PJ reconstruction: *rit. External cognate: Maxakalí /te, IRR te-t/ ‘to prepare’.
PSJ *θə/*θa-ŋ ‘to wash (clothes)’ > KGG *ɸa/ɸõ-ŋ*; LKL *ðo/ða-ŋ*
Possible PJ reconstructions: *jô(C), *jâ(C), *cô(C), or *câ(C). External cognate: Ofayé *xoh*.
PSJ *θu ‘vagina’ > KGG *ɸu*; LKL *ðu*
Possible PJ reconstructions: *ju or *cu. External cognate: Krenák *jo*, Ofayé *xa*.
PSJ *jur ‘to arrive’ > KGG *jur*; LKL *jul*
Possible PJ reconstruction: *jut. External cognate: Maxakalí /mũ-ca, IRR mũ-cã-t/ ‘to arrive/leave.PL’.
PSJ *ki ‘LOC’ > KGG, LKL *ki*
Possible PJ reconstruction: *ki(C). External cognate: Karajá *ki*.
PSJ *ŋgə ‘earth’ > KGG *ŋga*, LKL *ŋgo*
Possible PJ reconstruction: *ŋgô(C) or *ŋgâ(C). External cognates: Chiquitano *ky-*, Proto-Tupian *ƙɯc*.

5. Conclusions

Above I have exposed what I believe to be the first modern proposal regarding the reconstruction of Proto-Jê phonology and basic vocabulary.

Some of the ideas advanced in this paper may appear to be somewhat radical, such as the reconstruction of five contrastive vowel heights and of two coda types; these were, however,

the most parsimonious explanations I could find. It is nevertheless possible that an entirely different reconstruction will emerge in the future that will also account for the data.

Future research will have to tackle the issue of the unexpected nasalization in Central Jê, which remains completely unexplained in my current framework.

Another issue that remained beyond the scope of this paper is the reconstruction of Proto-Jê morphosyntax. It appears possible to project many morphosyntactic features shared by Northern Jê and non-Jê Macro-Jê languages, such as Maxakalí, onto the Proto-Jê level, even though a detailed study that would take into account Central Jê and Southern Jê data has not been carried out so far. These features include:

- AOV/SV, dependent-head constituent order;
- *split-S* alignment in clauses headed by finite verbs; ergative-absolutive alignment in clauses headed by non-finite verbs (cf. Castro Alves 2010, Nikulin & Silva to appear);
- obligatory expression of preposed internal arguments, leading to the occurrence of "expletive" third person markers in case of arguments dislocated from their canonical position;
- a clear-cut distinction between inflectable and non-inflectable nouns, whereby non-inflectable nouns must be preceded by a possessive classifier or a genitive postposition in order to be possessed, among other features.

Inflectional and derivational morphology of Proto-Jê also deserves to be explored in greater detail. Inflectional markers must have included person prefixes (encoding the possessor in nouns, the complement in postpositions, the O/S_O argument in finite verbs, and the O/S argument in non-finite verbs) and non-finiteness suffixes; derivational morphemes definitely included, but were hardly limited to, so called *formatives* (also known as *transitivity prefixes*, cf. Oliveira 2005, Salanova 2011b, Nikulin & Salanova forthc.). These topics should be covered in future research.

Data provenance

Panará	Bardagil-Mas 2018, Dourado 2001, Lapierre ms., Vasconcelos 2013
Southern Kayapó	Barbosa 1918
Xavánte	Estevam 2011, Hall et al. 1987, Lachnitt 1987
Xerénte	Krieger & Krieger 1994, Sousa Filho 2007, Souza 2008, Eneida Brupahi Xerente p. c., Mário André Coelho da Silva p. c.
Kaingáng	Wiesemann 2002
Laklãnõ (= Xokléng)	Alves Jr 2014, Bublitz 1994, Gakran 2015, Jolkesky & Gakran ms.
Ingain	Ambrosetti 1896 (*Pedro, Cosme Román, María Antonia*), Lista 1883 (*Lista*), Vogt 1904 (*Vogt¹, Vogt³*)

The PNJ reconstructions are based on data from numerous sources specified in Nikulin & Salanova (forthc.). For sources on Macro-Jê languages other than Jê, the reader is referred to Nikulin & Silva (forthc.).

References

Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2012. On Finnic long vowels, Samoyed vowel sequences, and Proto-Uralic **x*. In: Tiina Hyytiäinen, Tiina Jalava, Janne Saarikivi, Erika Sandman (eds.). *Per Urales ad Orientem. Iter polyphonicum multilingue. Festskrift tillägnad Juha Janhunen på hans sextioårsdag den 12 februari 2012*: 227–250. Helsinki: Suomalais-Ugrilainen Seura. (Suomalais-Ugrilaisen Seuran Toimituksia, 264.)

Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2015. The Finnic 'secondary *e*-stems' and Proto-Uralic vocalism. *Suomalais-Ugrilaisen Seuran Aikakauskirja* 95: 25–66.

Alves Jr, Ozias. 2014. *Parlons xokleng / laklãnõ. Langue indigène du sud du Brésil.* Paris: L'Harmattan.

Ambrosetti, Juan Bautista. 1896. Materiales para el estudio de las lenguas del grupo Kaingangue (Alto Paraná). *Boletín de la Academia Nacional de Ciencias de Córdoba* 14: 331–380.

Barbosa, Alexandre de Souza. 1918. *Cayapó e panará.* Uberaba: manuscript. Available online at: <http://etnolinguistica.wdfiles.com/local--files/biblio%3Abarbosa-1918-cayapo/barbosa_1918_cayapo.pdf>

Bardagil-Mas, Bernat. 2018. *Case and agreement in Panará (Naamval en congruentie in het Panará).* PhD. Groningen: Rijksuniversiteit Groningen.

Benware, Wilbur A. 1974 [1995]. *The Study of Indo-European Vocalism in the 19th century: From the beginnings to Whitney and Scherer.* Amsterdam, Philadelphia: John Benjamins. (Studies in the History of the Language Sciences, 3.)

Bishop, Nancy. 1996. A preliminary description of Kensiw (Maniq) phonology. *Mon-Khmer Studies* 25: 227–253.

Bublitz, Terezinha. 1994. *Análise fonológica preliminar da língua Xokléng.* MA thesis. Brasília: Universidade de Brasília.

Carvalho, Fernando O. de. 2016. On the Development of the Proto-Northern Jê Rhotic in Panará Historical Phonology. *Anthropological Linguistics* 58(1): 52–78.

Carvalho, Fernando O. de, Andrey Nikulin. In preparation. *Internal classification of Tupian.*

Castro Alves, Flávia de. 2010. Evolution of Alignment in Timbira. *International Journal of American Linguistics* 76(4): 439–475.

Constenla Umaña, Adolfo. 2000. La restitución: un método lingüístico reconstructivo sincrónico. *Filología y Lingüística* 26(2): 161–180.

D'Angelis, Wilmar da Rocha. 2002. Gênero em Kaingang? In: Ludoviko dos Santos, Ismael Pontes (orgs.). *Línguas Jê: estudos vários*: 215–242. Londrina: Eduel.

Davis, Irvine. 1966. Comparative Jê phonology. *Estudos Lingüísticos: Revista Brasileira de Lingüística Teórica e Aplicada* 1(2): 20–24.

Dourado, Luciana Gonçalves. 2001. *Aspectos morfossintáticos da língua Panará (Jê).* PhD. Campinas: Universidade Estadual de Campinas.

Estevam, Adriana Machado. 2011. *Morphosyntaxe du xavante. Langue jê du Mato Grosso (Brésil).* PhD. Paris: Université Paris Diderot (Paris 7).

Gakran, Nanblá. 2015. *Elementos fundamentais da gramática Laklãnõ.* PhD. Brasília: Universidade de Brasília.

Hall, Joan, Ruth Alice McLeod, Valerie Mitchell. 1987 [2004]. *Pequeno dicionário Xavante–Português, Português–Xavante. Damreme pibuzé hã a'uwẽ mreme, warazu mreme na te te ĩsaprĩ mono zém na duré warazu mreme, a'uwẽ mreme na te te ĩsaprĩ mono zém na.* Cuiabá: Sociedade Internacional de Lingüística.

Jolkesky, Marcelo Pinho de Valhery. 2010. *Reconstrução fonológica e lexical do Proto-Jê Meridional.* MA thesis. Campinas: Universidade Estadual de Campinas.

Jolkesky, Marcelo Pinho de Valhery, Nanblá Gakran. Ms. As classes verbais em Xokleng. Available online at: <https://www.academia.edu/5278237/As_classes_verbais_em_xokleng>.

Kassian, Alexei. 2015. Towards a formal genealogical classification of the Lezgian languages (North Caucasus): testing various phylogenetic methods on lexical data. *PLoS ONE* 10(2): e0116950.

Krieger, Wanda Braidotti, Guenther Carlos Krieger (orgs.). 1994. *Dicionário escolar: Xerente–Português; Português–Xerente.* Rio de Janeiro: Junta das Missões Nacionais da Convenção Batista Brasileira.

Lachnitt, Georg. 1987. *Romnhitsi'ubumro. A'uwẽ mreme = waradzu mreme. Dicionário xavante = português.* Campo Grande: Missão Salesiana de Mato Grosso.

Ladefoged, Peter, Ian Maddieson. 1990. Vowels of the world's languages. *Journal of Phonetics* 18: 93–122.

Lapierre, Myriam. Ms. *A phonological analysis of Panãra.* Berkeley: University of California.

Lista, Ramón. 1883. *El Territorio de Las Misiones*. Buenos Aires: Imprenta «La Universidad», de J. N. Klingelfuss.

Nikulin, Andrey. 2015. Apofonia e sistema vocálico do Proto-Jê Meridional: contribuição para estudos comparativos das línguas Jê. *Revista Brasileira de Linguística Antropológica* 7(2): 275–306.

Nikulin, Andrey. 2016. Historical phonology of Proto-Northern Jê. *Journal of Language Relationship* 14(3): 165–185.

Nikulin, Andrey. 2017. A phonological reconstruction of Proto-Cerrado (Jê family). *Journal of Language Relationship* 15(3): 147–180.

Nikulin, Andrey, Andrés Pablo Salanova. Forthc. Northern Jê verb morphology and the reconstruction of finiteness alternations. *International Journal of American Linguistics* 85(4).

Nikulin, Andrey, Mário André Coelho da Silva. Forthc. As línguas Maxakalí e Krenák dentro do tronco Macro-Jê. *Cadernos de Etnolingüística (Estudos de Lingüística Sul-Americana)* 7.

Nikulin, Andrey, Mário André Coelho da Silva. Forthc. A evolução do alinhamento na língua Maxakalí. In: Christiane Cunha de Oliveira (org.). *Memórias do II Encontro dos Americanistas no Cerrado*. Goiânia: Universidade Federal de Goiás.

Oliveira, Christiane Cunha de. 2005. *The language of the Apinajé people of Central Brazil.* PhD. Eugene: University of Oregon.

Pache, Matthias. 2018. *Contributions to Chibchan Historical Linguistics.* PhD. Leiden: Universiteit Leiden.

Ribeiro, Eduardo Rivail. 2004. Prefixos relacionais em Jê e Karajá: um estudo histórico-comparativo. *Línguas Indígenas Americanas* 4: 91–101.

Ribeiro, Eduardo Rivail. 2011. Prefixos relacionais como evidência histórico-comparativa: os casos Chiquitano e Jabutí. In: Ana Suelly Arruda Câmara Cabral, Aryon Dall'Igna Rodrigues, Jorge Domingues Lopes, Maria Risolêta Silva Julião (orgs.). *Línguas e Culturas Macro-Jê, vol. 2*: 105–120. Campinas: Curt Nimuendajú; Brasília: Laboratório de Línguas Indígenas.

Ribeiro, Eduardo Rivail. 2012. *A grammar of Karajá.* PhD. Chicago: University of Chicago.

Ribeiro, Eduardo Rivail, Hein van der Voort. 2010. Nimuendajú was right: The inclusion of the Jabutí language family in the Macro-Jê stock. *International Journal of American Linguistics* 76(4): 517–570.

Salanova, Andrés Pablo. 2011a. A flexão de terceira pessoa nas línguas Jê. *Línguas Indígenas Americanas* 11: 75–114.

Salanova, Andrés Pablo. 2011b. Reduplication and number in Mẽbengokre. In: Gale Goodwin Gómez, Hein van der Voort (eds.). *Reduplication in Indigenous Languages of South America*: 252–273. Leiden: Brill.

Sousa Filho, Sinval Martins. 2007. *Aspectos morfossintáticos da língua Akwẽ-Xerente (Jê).* PhD. Goiânia: Universidade Federal de Goiás.

Souza, Shelton Lima de. 2008. *Descrição fonético-fonológica da língua akwen-xerente.* MA thesis. Brasília: Universidade de Brasília.

Traunmüller, Hartmut. 1982. Der Vokalismus im Ostmittelbairischen. *Zeitschrift für Dia-lektologie und Linguistik* 49: 289–333.

Vasconcelos, Eduardo Alves. 2013. *Investigando a hipótese Cayapó do Sul-Panará.* PhD. Campinas: Universidade Estadual de Campinas.

Vogt, Federico. 1904. Die Indianer des Obern Paraná: II. Die Indianer der Coroados-Gruppe. *Mitteilungen der Anthropologischen Gessellschaft in Wien* 34(4): 353–377.

Wiesemann, Ursula. 1978. Os dialetos da língua Kaingáng e o Xokléng. *Arquivos de Anatomia e Antropologia* 3: 199–217.

Wiesemann, Ursula Gojtéj. 2002. *Dicionário Kaingang–Português, Português–Kaingang.* Curitiba: Esperança.

А. В. Никулин. Реконструкция фонологии и словарного состава пра-же языка

В статье рассматриваются регулярные фонетические соответствия между праязыками нагорной (Никулин 2017) и южной (Жолкески 2010) ветвями языковой группы же, самой глубокой генетической единицы в составе южноамериканской семьи макро-же. На основании этих соответствий предлагается реконструкция фонологии праязыка семьи же. Автор восстанавливает 11 согласных и 19 гласных фонем. Максимальная структура слога, восстанавливаемая для пра-же, — */CrVC/; при этом имеется ряд ограничений на сложные инициали (восстанавливаются только */pr, mr, kr, ŋr/). Автор восстанавливает правило, согласно которому носовые инициали имели в пра-же постоализованные аллофоны перед неносовыми централями. Статья завершается списком этимологий же.

Ключевые слова: языки же, языки макро-же, языковая реконструкция, сравнительно-исторический метод.

Giorgio ORLANDI
independent researcher; thegaygenius@gmail.com

On the classification of the Ng Yap dialects: some thoughts on the subgrouping of Sinitic languages*

The Ng Yap (formerly Sze Yap) dialects are routinely considered a branch of the Yue subfamily. This paper seeks to demonstrate that, contrary to this widespread opinion, these dialects show a wide range of distinctive features which, for formal purposes of language/dialect classification, may warrant their separation from the Yue subfamily. This paper also discusses the criteria which are often at the basis of language subgrouping in the field of Chinese linguistics. Nevertheless, this work should be regarded only as an attempt of stimulating a further discussion into a topic which has been overlooked for far too long.

Keywords: Ng Yap dialects, Yue dialects, language subgrouping, Sinitic languages, Chinese dialects.

«The chief danger to our philosophy, apart from laziness and woolliness,
is scholasticism, which is treating what is vague as if it were precise
and trying to fit it into an exact logical category»
Frank Pulmpton Ramsey, *Philosophical Papers*, 1929, p. 269.

1. Introduction

The Ng Yap dialects, formerly known as Sze Yap (or Seiyap), are spoken primarily in the Guǎngdōng and Guǎngxī provinces, as well as in Macao, Hong Kong and in many overseas communities. They are generally considered a branch of the larger Yue subfamily, and may be further divided geographically into two sub-branches, Xīn'ēn 新恩 and Kāihè 開鶴. The most prestigious and perhaps representative Ng Yap dialect[1] is Hoishanese ([hɔiꜜsanꜜwaꜜꜜ] or Toishanese).

* This paper incorporates elements of an unpublished draft presented during the 23rd International Symposium on Yue Dialects, which was held on 15th and 16th December 2018 at Jìnán University (Guǎngzhōu, PRC). In some cases, the present writer has followed the advices received; in other cases, this author has, instead, decided to follow his way, perhaps at his own peril. In general, the presentation has met with enthusiasm, though some specialists have stipulated that some revisions, which shall be illustrated and answered in the present paper, must be done.

[1] In the present paper terms such as 'language' and 'dialect' are basically interchangeable. The distinction is somewhat arbitrary, and, at times, based on criteria (e.g. prestige, correctness, *etc*.) which only show a certain degree of linguistic unsophistication (Trask, 2007: 49–50). In China, Mandarin Chinese is the standard language, which is to say that it is the codified variety superimposed over regional dialects, and typically used in formal settings and for education. Nevertheless, it is important to stress that no such thing as modern standard Chinese (*pǔtōnghuà* 普通話) existed, say, 200 or 300 years ago. Only recently a bunch of social and especially political measures have combined to give rise to a variety of Mandarin, also referred to as *guóyǔ* 國語 'national language,' accepted as the standard language in the whole country (Norman 1988: 135–137). In other words, we may agree with Max Weinreich in believing that "a shprakh iz a dialekt mit an armey un flot."

Descriptions and sources of information about the Ng Yap dialects are not abundant, even though they have a long history. One of the first mentions of a Ng Yap dialect (a variety called Llin-nen or Hsin-ning) can be found in Wells Williams's *A Tonic Dictionary of the Chinese Language in the Canton Dialect*, where it is stated that:

> The people from the district of Sinhwui 新會 lying south-westerly from Canton, exhibit the most remarkable peculiarities in pronunciation, and it is a puzzle to the scholars in the city how they should have originated. (Williams, *op. cit.* 1856: ix).

Of much greater interest is the section *Tǔyán* 土言 (local speeches) of the *Guǎngdōng xīnyǔ* 廣東新語, a work written by the scholar Qū Dàjūn 屈大均 (1630–1696), where we find scraps of information about the Xīnhuì dialect. For example, in this work it is mentioned that the eldest son is called *dziaŋ*^{55} *dɔi*^{55} 長仔 or $^{m}b^{w}ɔn$ (?) 屘 in Xīnhuì. It is curious to note that this last form is commonly used in Southern Mǐn (*ban*^{1}) to indicate the youngest son. For instance, it was recorded in both the *Lūi-im Biāu-gō͘* 彙音妙悟 (1800) and the *Cheng-pó͘ Lūi-im* 增補彙集妙悟 (1820)[2]. In the former, it is assigned to the *dān* rime 丹韻 (*-an), with *wén* initial 文母 (*b-) and marked by an upper even tone; in the latter, it is instead described as belonging to the *gàn* rime 干韻 (*-an) with the *mén* initial 門母 (*b-). It is absent in Douglas' *Chinese-English Dictionary of the Vernacular or Spoken Language of Amoy* (1873), but it can be found in the *Supplement* added by Rev. Thomas Barclay (1923). It is also found in Ogawa's dictionary (1908: 551–553).

Returning to Ng Yap dialects, the first real description of the language is attributable to the New Zealand Presbyterian missionary Alexander Don (1857–1934).[3] In his two papers dedicated to the "Llin-nen variation of Cantonese," Don describes quite accurately the phonological system and the tonal behaviour[4] of the Xīnníng dialect, with an eye towards comparisons with standard Cantonese, as described by Williams (1812–1884) and Parker (1849–1926). Like Williams, Don is silent on the position of Xīnníng and on the internal structure of the Yue subfamily in general.

Other studies have been dedicated to the study and the description of Ng Yap. Among these, we may cite Yiu (1946), Chao (1951), Cheng (1973), Him (1980), Light (1986), Lee (1987), Tong (1997), Yue-Hashimoto (2006), Kwok (2006), Takekoshi (2017), *etc*. Few studies, however, are concerned with the classification of the Ng Yap dialects. Nevertheless, before analysing and discussing subgrouping within the Yue subfamily, the discussion would benefit from a paragraph dedicated to subgrouping within the Sinitic family in general.

2. Dialect classification and subgrouping: what should subgrouping be based on?

First a few questions: out of a group of languages, if two or more languages are more similar to each other, can we safely hypothesise that they are languages of the same group or family? *Pace* anyone who would answer 'yes,' the correct answer is actually 'no.' As any good textbook in historical linguistics has demonstrated, with ample and documented examples, overall similarities by no means imply genetic relationship or affiliation, because there are many reasons why languages exhibit similar characteristics, and relationship is not necessarily

[2] The best treatment of this work is Ang Uijin (Hóng Wéirén) 洪惟仁. *Huìyīn miàowù yǔ gǔdài Quánzhōu yīn* 彙音妙悟與古代泉州音 [The Lūi-im Biāu-gō͘ 彙音妙悟 and the sound system of the ancient Quánzhōu language]. Guólì zhōngyāng túshūguǎn Táiwān fēnguǎn 國立中央圖書館臺灣分館, 1996.

[3] See, for instance, Don 1883 and Don 1884.

[4] In the *Guǎngdōng xīnyǔ* it is mentioned that the Xīnhuì dialect often merges *píng* tones with *zè* tones (*Xīnhuì yīn duō yǐ píngzè xiāng yì* 新會音多以平仄相易).

one of them. What if two (or more) languages, very close each other, exhibit a common feature? Is it safe to suppose that those languages are related or members of the same family? Surprisingly, still not necessarily. For languages sometimes may exhibit similar or even identical characteristics not because they were inherited by from a common ancestor language, but because of parallel developments, just like Hakka resembles common Mǐn in that the upper and lower entering tones have switched their places (i.e. the so-called lower entering tone in fact is higher in pitch than the upper entering tone), but this is, in all probability, a parallel development. In evolutionary biology, the former (i.e. an homologous feature) is called homology; the latter (parallel developments) homoplasy. Not differently from biology, linguistics—and it is hoped that the reader will forgive the paternalistic tone of this statement—establishes language classification only on the basis of "synapomorphies," namely on "recent" shared homologies (i.e. shared innovations), because only shared innovations may prove that a closely relation is likely to exist between two or more languages, or may successfully give us information about phylogenies[5]. Thus, it is up to the competent linguist to ferret out common apomorphic (derived) characters, distinguishing them from plesiomorphic (primitive) features and from parallel developments.

2.1. The subgrouping of the Sinitic family

It seems that dialect classifications within the Sinitic family have generally relied on the phonemic features of medieval Chinese (Wang 1936, *cf.* Wang 1996: 249, Li 1937: 1–13, Ting 1982: 258).[6] According to Li Fang-kuei (1937, 1938), the treatment of medieval Chinese voiced and aspirated stops is a crucial feature for the subgrouping of Chinese dialects. Ting Pang-hsin (1982) suggested to separate "early historical features" (*zǎoqi lìshǐxìng de tiáojiàn* 早期歷史性的條件) from "late historical features" (*wǎnqi lìshǐxìng de tiáojiàn* 晚期歷史性的條件), although it is not very easy to understand what is meant by these two terms, and how to distinguish them. Norman (1988: 182) has proposed a classification according to phonological, grammatical and lexical items. Lau (2002: 82) has proposed a new classification which takes into consideration a feature which was apparently overlooked by most scholars, i.e. the sound change according to which the "voiced rising tone" (*zhuó shǎng* 濁上) becomes "voiced departing" (*zhuó qù* 濁去)[7]. One of the most valuable approaches, perhaps, is Simmons (1999), but he is apparently concerned only with Hángzhōu and Northern Wú in general[8].

It seems to the present writer that it is necessary to remark that the phonological approaches mentioned above represent, more often than not, more an attempt to demonstrate how a given Sinitic language or group of languages have evolved, in a somewhat mechanistic

[5] In biology, the scientific methodology which groups organisms on the basis of derived shared characteristics is called 'phylogenetic systematics' (also known as cladistics). The trend of grouping organisms which share derived features was apparently started by the German entomologist Willi Hennig, see Hennig 1950.

[6] In the opinion of the present writer, terms such 'early Chinese' or 'medieval Chinese' are more advantageous, so long as they are supposed to indicate a *Spracheinheit*, intended as an abstraction of linguistic (in this case mainly phonemic) features that may have been common to a given group of speakers, at some time or other, and not a living *koine*.

[7] For further knowledge about the sound change involving the voiced rising tone, see also Ho 1988.

[8] However, it seems that a series of related problems have pushed many scholars to adopt different approaches. Among these, we may mention the computational and statistical approaches that have been used by Zhōu Zhènhè 周振鶴 and Yóu Rǔjié 游汝杰 (1985), or by Cheng Chin-chuen 鄭錦全 (1987, 1994, 1996). Nevertheless, they suffer from serious linguistic flaws (see, e.g., Yue-Hashimoto 1991: 165), and will not be discussed in detail here. A similar, though much better, approach is suggested in Baxter 2006.

fashion, from the sound classes of medieval Chinese (if any), or an attempt to quantify how many medieval Chinese features have been preserved by modern dialects, than an analysis which focuses exclusively on those dialects and on their sound systems. This must be certainly misleading, for if the sound classes of medieval Chinese are used as a phlogiston which act as a constraint on linguistic reconstruction or classification, rather than acting as a device to broaden the horizon of possibilities, then they render the identification or reconstruction of deviant features or of different sound changes impossible a priori, as they introduce a bias toward what is frequent and regular in the sound system of rime tables. This should not be taken to imply that all the analyses which are based on medieval Chinese sound classes must be necessarily wrong, but the great confidence these scholars, or at least part of them, have towards medieval Chinese, as if it were a real language and not a diasystem which is the product of our theoretical abstractions, strikes the present writer as incredible.

In addition, it seems that the field of Chinese linguistics is also plagued with the widespread and undemonstrated belief, according to which unwritten features or deviant features must not be ancient, or must be derived from a sole common source (e.g. a given sound class of medieval Chinese). Many scholars also claim that the comparative method cannot be applied within the Sinitic family (Hirata 1988, Wáng 1999, 2004, 2005), because the monosyllabic structure of the Chinese morpheme, and the extensive borrowing throughout the Chinese speaking area render infeasible its application, ignoring the fact that the comparative method, which is independent of "lexical typology," is exactly a tool for eliminating chance resemblance, universals, and borrowings as plausible causes for cross-linguistic similarity.

All these prejudices may be understandable in view of the inevitable training to which the general Chinese historical linguist, including the present writer, is routinely submitted, but we must realise, once and for all, that we are in a more advantageous position, and thus we do not need to justify or accept our imprecise—and at times even grotesque—terminology/approach by tracing them back to the Míng (1368–1644) and Qīng (1644–1912) philological traditions[9].

Unfortunately, many scholars who have recognised the limits of the philological-phonological approach, instead of working within the framework of widely accepted, recommended practices of historical linguistics, have preferred to resort to "mutual intelligibility" as the main criterion for subgrouping, since structural methodologies based on phonological characteristics have been considered "too complex" (Tang 2017: 553). However, this writer is inclined to question mutual intelligibility as a criterion for language subgrouping, otherwise Spanish might well be considered a dialect of Italian, while Bergamasque should be regarded, instead, as a separate, distinct Romance language. In other words, shared innovations, instead of shared retentions (i.e. the retention of features from a common source, in this case the sound system of medieval Chinese) and mutual intelligibility, should be the principal criterion for subgrouping.

2.2. The subgrouping of Yue

The first attempt of subgrouping within the Yue subfamily has been made by Zhān Bóhuì (1981). Nevertheless, it seems that this attempt was prevalently aimed at validating an earlier

[9] It seems, however, that younger scholars are actively applying the 'comparative method' to gain further information about the morphophonological system of the various Chinese dialects. See, for instance, Chén Ruìqīng (2018), Wu Rui-wen (2014), Zhāng Jìngfēn (2013), *etc*. However, they are not concerned with subgrouping, and, in fact, are not entirely liberated from the post-Karlgrenian tradition which they apparently wish to reject. This, of course, holds true in part also for the present paper, but if the aim of this paper is to convince its critics, then it should use arguments that most of them would accept.

proposal by Yuán Jiāhuá (1960). More worth discussing is the classification of the various Yue languages spoken mainly in the Guǎngxī province by Yáng Huàndiǎn (1985). Yáng has recognised four branches within Guǎngxī Yuèyǔ, *viz.* Guǎngfǔ (Wúzhōu, Cāngwú, Hèxiàn, Dānzhú, Dà'ān), Yōngxún (Nánníng, Yángzhōu, Yōngníng, Chóngzuǒ, Níngmíng, Héngxiàn, Guìpíng, Píngnán, *etc.*), Gōulòu (Yùlín, Běiliú, Róngxiàn, Língxī, Téngxiàn, Méngshān), and Qīnlián (Qīnzhōu, Hépǔ, Liánzhōu, Língshān).

Xióng Zhènghuī (1987) represents the first attempt of subgrouping based on linguistic features. In his view, the most important feature was the treatment of medieval Chinese 'entirely muddy' initials (*quán zhuó yīn* 全濁音). He realised that medieval Chinese voiced initials had, in part, become voiceless aspirated; in part, they had become tenuis. To the former belongs the Wúhuà branch, which comprises the dialects of Wúzhōu, Huàzhōu and Zhànjiāng; to the latter belongs the Gōulòu branch, which includes the dialects of Sìhuì, Guǎngníng, Déqìng, Luódìng, Yùnán, Fēngkāi, Huáijí, Yángshān and Liánshān. Other three branches were recognised by Xióng on the basis of certain phonological features. For example, they all agree in having aspiration only with level and rising tones, not with entering and departing. Sìyì (Sze Yap) dialects (Táishān, Ēnpíng, Hèshān, Xīnhuì, Jiāngmén, Dòumén, Kāipíng) agree in having [h] for traditional *tòu* initials 透母 /*tʰ/. The Gāoyáng branch, which includes the dialects of Yángjiāng, Yángchūn and Gāozhōu, differs from Guǎngfǔ in showing a voiceless alveolar lateral fricative [ɬ] for traditional *xīn* initials 心母 /*s/.

A more recent, and more valuable subgrouping is provided in Yue-Hashimoto (2006). However, she relied mainly on mutual intelligibility, a criterion that the present writer is inclined to question.

If we exclude Yue-Hashimoto (2006), all other classifications have been based on both shared innovations and shared retentions, as dialects have been drawn close to each other both when they retained a feature of medieval Chinese (e.g. aspiration, voicing, *etc.*) or when they showed a shared innovation (e.g. when showing different phonemes instead of the expected ones, according to the sound classes of medieval Chinese). This must be only partially correct, since classification based on shared retentions is not a recommended practice in historical linguistics. Nevertheless, it is in the opinion of the present writer that these attempts are certainly valid, although better and more reliable results might be obtained by looking at those innovative features in morphology, phonemics and lexicon. As the subgrouping of the entire Yue subfamily is abundantly beyond the scope of the present paper, the analysis shall focus only on the Ng Yap branch.

2.3. In defence of the *Stammbaum* model

Since subgrouping often involves the internal classification of languages within a family, subfamily or group, normally represented in a family tree, it seems necessary to spend a few words defending the *Stammbaum* model, whose dismiss, it seems, has become fashionable in these years, especially in the field of Chinese linguistics. This writer is aware of the fact that network models are also recommended, especially when dialects are very close in space and in time, when there has been extensive borrowing between two or more languages, when the speciation of languages is never proceeded by an abrupt separation, or when each internal node of the alleged tree is constrained to represent virtually undifferentiated dialects (Ringe et al. 2002: 106). However, contrary to what many authors believe, the two methods are not mutually exclusive. Recently, network models for the subgrouping of Sinitic have been used by Zhang et al. (2018). Their approach, i.e. an admixture inference to decompose the underlining structure of the diversity of Sinitic languages based on phonemic inventories, is laudable but

does contain some peculiarities. They reject the *Stammbaum* theory, but feel safe to rely on migrations and on historical socio-genetic speculations. Alas, they do not provide illustrative examples and counter-examples, or make suggestions of what should be done with such language families as Sinitic where Tree models — they say — don't work. The relationship between languages/dialects and socio-genetic history is misleading: language is independent of genes, because social groups change their languages for different reasons and at different epochs. Nevertheless, Zhang et al. (2018: 4) state that Tree models are useless in Sinitic and only network models can work. This author insists that it is not fair, nor historically accurate, to call the entire Tree model theory into question on the grounds that it is based on an unrealistic concept or on an unsuitable scenario, especially when this claim is not discussed in detail, but is plagued with a partial misunderstanding of the Tree model itself, of its application, and of its finalities. Furthermore, since most of the old Sinitic linguistic territory is unknown, the *Wellentheorie* model cannot be applied in any meaningful way to determine dialectal relationship, at least not as Zhang et al. believe. There have been various attempts to determine how many "dialects" were spoken in Ancient China and to which modern, received languages they correspond. Some scholars even claimed to have discovered the relationship between some of the dialects of the Warring States period (475–221) and their received daughter languages (*cf.* Xǔ Wénxiàn 2001; Zhào Tóng, 2006; Hú Hǎiqióng, 2012). This writer may be mistaken, but what he sees here is an amazing lack of understanding of the linguistic history of what is present-day China (but back in those times was not) and the survivorship bias. We know that, prior to Cristoforo Colombo's first voyage in 1492, there were presumably twice as many languages as there are today. It is not difficult to imagine that, projecting backwards through time, there have reasonably existed thousands and thousands of languages, many of them also on present-day Chinese territory.[10] Therefore, in the absence of a clear and detailed description of the languages of the Warring States period (which is lacking), we cannot affirm which was the language of the reign/chiefdom X, Y or Z, and of which received language they were the alleged ancestors. This makes unknown most of the old linguistic territory of present-day China, *cum bona pace* of anyone who thinks otherwise. Hence, we are left with the *Stammbaum* model and the hypothesis according to which the various Sinitic groups have gradually dispersed, in a more or less unknown order (Mǐn dialects are typically considered to have split before other dialects, rightly in this writer's opinion, but the split order of other groups is not easy to detect), from a common ancestor (medieval Chinese? Early Chinese?) which can and has been reconstructed by means of a methodology, which, with all its limits and difficulties, has been continuously refined over the years.

Both Tree and Network models are discussed in Mahé and Wang (2006). Mahé and Wang also claim that the tree model is not suitable for Sinitic, because cases of diglossia are attested since ancient times, which is true in principle but only partly so. Diglossia was attested in other parts of the world just as well as in China: Greek, for instance, was spoken in Asia Minor and Southern Italy along with many other languages. Furthermore, it is not clear why they are so sceptical towards the *Stammbaum*, a method which they seem to misunderstand, and yet feel safe to rely so heavily on lexical items, in spite of the extensive borrowing which has occurred throughout the Sinitic-speaking area. In addition, in Mahé and Wang 2006, trees are rooted using Old Chinese, a language never well defined but always imagined in linguistic terms drawn from later periods, and reconstructed on the basis of sources of evidence which are far from being contemporary with each other. This writer does not seek to take issue with

[10] Authors such as Bickel (2014: 120, n.5) have calculated that there may have been half a million languages around 100k years ago, based on current rates of stability of languages and of language death.

the irenic spirit of their work, but it is necessary to voice one's disagreements over misconceptions in a linguistic debate, no matter how laudable one's work might be (and their work certainly is).

The *Stammbaum*, as a linguistic concept, is not an absolute, and it has been revised and improved many times in its long history. Yet, many critics treat it as if it were a universally-applicable method, static and immutable in time. In fact, as rightly pointed out by Rasmussen (1991: 467), the Tree model theory does not have the scope of telling us how linguistic unity came about or fell apart, but simply to inform us about the existence of unity and disunity of a group of languages. It does *not* postulate a past without variations, nor does it presuppose a lack of variation[11]. Thus, contrary to the opinion of many dialectologists, the *Stammbaum* theory may and does work within the framework of Sinitic (the testimony of other models, nevertheless, may yield even better results), and the huge autonomy of its individual branches plainly reflects the clear-cut splits of the Tree model, further validating both the *Stammbaum* model per se and its applicability within the Sinitic family. It follows that a subgrouping of the Yue subfamily or of the Ng Yap dialects based on the Tree model is perfectly reliable.

3. Peculiarities and innovations of the Ng Yap dialects

As discussed in section 2.2, it seems that there has always been, in the field of Cantonese linguistics, the general and implicit assumption that Ng Yap must not be a separate branch, despite all the striking divergences it shows.

To the best of this writer's knowledge, the general methodological procedure of stating that a language X is not part of the family Y would be: (a) list certain forms or features that are commonly diagnostic of all or most dialects of a given family; (b) show that a language does not belong to that family because it lacks most of these diagnostic forms and features, preventing it from being classified as such. Nevertheless, it seems that Yue dialects cover a wide range of diverging features, so it is difficult to know what Yue really is, and, in fact, the present writer suspects that Yue is instead a false taxon. For instance, a particularity of Yue dialects (excluding Ng Yap, Yángjiāng and Yángchūn) is their treatment of medieval Chinese 'entirely muddy' affricates as voiceless unaspirated consonants, and yet many exceptions can be found in Nánhǎi Jiǔjiāng, Fóshān, Gāomíng, Sānshuǐ, and other Yue dialects. The lower even tone merges with the lower departing tone in the dialects of Cāngwú and Guìpíng, but merges with the upper departing tone in the dialects of Shùndé and Nánhǎi. Given that no systematic reconstruction of proto-Yue has been done, and that we do not know whether those unusual phonemic features are retentions or innovations, this methodological procedure is probably excluded a priori, and we can only assume that Ng Yap dialects should be regarded as a distinct branch, because they do not behave like all other varieties of Yue are supposed to behave, especially in the treatment of initials and tones. Hence, we can use a sort of "*apophatic taxonomy*" to reach the conclusion according to which Ng Yap dialects are not Yue, by accumulating a list of features which are absent in any other variety of Yue besides Ng Yap.

The first scholar who systematically used Sze Yap data to gain further knowledge about the Yue subfamily and to reconstruct certain aspects of the sound system of proto-Yue is

[11] The very concept of 'variation' is somewhat misleading. Of course, for a discipline such as dialectology, which is all about variation, this concept is inevitably maximised, but general experience tells us that "variation is generally short-lived and territorially restricted" (Rasmussen 1991: 464).

McCoy (1966). Nevertheless, it is quite regrettable that McCoy's work showed no hint of interest in such a recommended practice of historical linguistics as internal reconstruction (rather, his reconstruction is fundamentally an assessment of the sound classes of these dialects on the basis of the ones found in rime tables), which may be extremely useful for exploiting many important data available since the end of the nineteenth century, when the systematic study of Ng Yap took place under the hands of Alexander Don (1883). It is likewise regrettable, in the opinion of the present writer, that McCoy (and others) did not show any knowledge of (or interest for) lexicostatistics, a method which is certainly not a recommended practice (with notable exceptions) in historical linguistics, but that, with all its uncertainties and shortcomings, could still be as good a way for dealing with data which are not easily verifiable.[12]

Generally speaking, the Ng Yap dialects show a wide range of features and innovations which are not shared by standard Cantonese and other Yue dialects. Apart from the set of correspondences argued in the previous sections, Hoishanese, Ēnpíng, *etc.* have prenasalised stops /ᵐb/ /ᵐd/ /ᵐg/. This peculiar feature appears to be a recent innovation, since there is no trace of it in Don's article. Prenasalised stops are also found elsewhere within the Yue subfamily, though they show some differences[13]. But Ng Yap, especially Hoishanese, also shows another type of nasalisation, which was recorded by Don, and that still occurs today, especially in those words which, according to traditional terminology, are classified as *yǐ* initials (*yǐmǔ* 以母):

In the version described by Don (1883), 'word' and 'play' are homophonous, while today we can observe that assimilation of the precedent velar nasal has occurred. This may lead us to talk about some of the many lenitions and fortitions that occurred in Hoishanese. Many words that originally had a velar nasal in initial position have changed their initial into a plain voiced velar plosive /g/, followed by the formation of a vocalic diphthong. Other varieties of Ng Yap, such as the Hèshān dialect, show a lenition of the bilabial plosive /p/ to voiced labiodental

[12] This is not to be taken to imply that the present writer recommends lexicostatistics and glottochronology, two methodologies of dating which are in fact partly independent of each other, as a way to establish genetic relationship. It is true that most linguists reject lexicostatistics and glottochronology (Campbell & Poser, 2008: 303 footnote), but to interpret them as two tools for establishing genetic relationship is neither fair to those scholars, such as Rafinesque and Broca, who contributed to invent them, nor historically accurate. Both lexicostatistics and glottochronology do not involve questions of proof, but can be pursued entirely within the frameworks of accepted linguistic families. In this specific case, they might be a useful tool to help establish a certain course for the study of the Yue family as a whole, and not to establish whether Ng Yap dialects do or do not belong to the Yue subfamily. Lexicostatistics has been clearly remodelled on the basis of radiometric dating, a scientific method which has proved successful in other fields of science. However, just like radiometric dating works only on rocks which cool from a liquid melt, such as basalts or granites, both of which solidify from lava, and not on fossils which were formed from dumped sediments, in the same way both lexicostatistics and glottochronology are not universally applicable practices. But to deny that they may have their advantages is a violence to the history of these methods and to all the serious practitioners who have helped to create, develop and adjust these techniques. For a lexicostatistical attempt to estimate the time depths of five major Sinitic languages, see Wáng Yùdé 1960; however, see Matisoff 2000 for a rejection of glottochronology in Sino-Tibetan linguistics in general.

[13] For this reason, scholars such as Ting Pang-hsin & Zhāng Shuāngqìng (2002: 207), or Liú Xīnzhōng (2010, personal communication) distinguish *bí guān sèyīn* 鼻冠塞音 (nasalised stops) from *hòu sè bíyīn* 後塞鼻音 (prenasalised stops). Although the present writer does not agree wholeheartedly with this terminology, he thinks that they are right in pointing out that a difference likely exists: in the case of Ng Yap we observe the articulation of a plosive segment which is realised with a brief period of air flow through the nasal cavity; in other cases, we observe a phonemic process where a segment, which does not involve oral closure (with consequent lowering of the velum), acquires nasalisation.

Table 1. Comparative table of Hoishanese words in traditional *yǐ* initials. Hoishanese pronunciations have been taken, with minor revisions, from Dèng Jūn 2006.

character	English gloss	Don's notation	pronunciation [IPA]
月[14]	moon	ngut	[ŋgut^{32}]
日[15]	day	ngit	[ŋgit^{21}]
言	word	ngun	[ŋgun^{21}]
玩	play	ngun	[gᵘɔn^{32}]
雅	refined	nga	[ŋga^{21}]
牛	cow	ngeu	[ŋgeu^{11}]
我[16]	I	ngoe	[gᵘɔ^{21}]
眼	eye	ngan	[ŋgan^{55}]
魚[17]	fish	*absent*	[ŋgui^{11}]
外[18]	outer	*absent*	[ŋgai^{32}]

fricative /v/, e.g., *pɛk[19] > *pak* > *viak* 'one hundred.'[20] Traditional *pāng* initials 滂母 /*pʰ-/ have become /h/ in both Hèshān and Kāipíng.

In most of Ng Yap dialects, the palatal approximant /j/ is an allophone of /ʒ/, but can also be an allophone of the close front vowel /i/ when used as a glide. Similarly, /w/ can be an allophone of the vowel /u/. The palatal sibilants are allophones of their equivalent alveolar sibilants in cases such as when the first vowel of the final consonant is a vowel which, according to the Jakobson-Halle distinctive feature system (1956), either is acute and non-flat /i/ or is characterised by a low second formant /u/. Like standard Cantonese, Hoishanese and other Ng Yap dialects aspirate in the lower rising tone and in the even tone, and routinely confine occlusive initials in the lower rising tones, but unlike Cantonese they do not develop aspirate stops into fricatives (Cantonese shows instead a marked predilection for fricatives over aspirates, even though the voiceless aspirated occlusive generally remains so in lower level and rising tones). The treatment of coronal sibilant /*s/ and postalveolar sibilant /*ʂ/ is very peculiar. These sibilant phonemes have presumably become a voiceless alveolar lateral fricative /ɬ/, as in Welsh.[21] However, this phoneme appears to be an areal feature: some linguists, such as

[14] In Hoishanese the word for 'moon' can be pronounced in the *shàng yáng rù* (high lower entering) tone [ŋgut^{32}] as well as in the *xià yáng rù* (low lower entering) tone [ŋgut^{21}].

[15] The word *ngit* 'day' in modern Cantonese can be found only in the *xià rù* tone, while in the Hoishanese version described by Don it could be found in *shàng rù*, *zhōng rù* and *xià xià rù* tones as well (Don 1884: 479).

[16] The 1st person pronoun can be pronounced in the *yīn píng* tone as [gᵘɔ^{33}] (literal reading) and [gᵘɔi^{33}] (colloquial reading) and in the *yáng shǎng* tone as [gᵘɔi^{21}].

[17] The literal reading of 'fish' is [ŋgui^{11}], while its colloquial reading is [ŋgui^{21}]. 'Fish' can also be pronounced [ŋgui$^{11\text{-}55}$] or [ŋgui$^{21\text{-}55}$] as a result of changed tone.

[18] 'Outer' deserves a special mention. It is a *yáng qù* word which is pronounced as [ŋgai^{32}] (and which sometimes exhibits a changed tone phenomenon, [ŋgai$^{32\text{-}55}$]) or as [gᵘɔi^{32}], with assimilation of the velar nasal.

[19] This "proto-Ng Yap" form has been reconstructed by the present author.

[20] In Hoishanese the phoneme /v/ may have evolved from an ancient *w, see Lau 2007: 169–74. For further information about lenitions in Ng Yap dialects, see Zēng 2014: 96–104.

[21] In his monumental *Études* (Chap. VI), the great Swedish sinologist Bernhard Karlgren doubted about the existence of this phoneme: "[l]atérale dentale, orale, sourde, le 'll' du dial. celtique du pays de Galles, par ex. dans Llewellyn, existerait selon M. A. Don (China Review, Vol. XI) dans le parler de Sin-ning du groupe Yue, comme représentant d'un ancien s, renseignement qu'il faudra vérifier." For further reading, see Karlgren 1915–1926: 270.

Lǐ Jǐnfāng (2002) and Mài Yún (2010), attribute the distribution of /ɬ/ to a Kradai substratum, even though others (*cf.* de Sousa 2015) find this scenario problematic. This phoneme, however, is found also in Southern Gàn and in other dialects of North Fújiàn. The change from *s to ɬ is typologically unusual, but does have some parallels. Most Central Tai and Northern Tai languages show this sound change (probably via an intermediate stage of n̥, *viz.* *s > n̥ > ɬ). In addition, Arapaho exhibits the extremely unexpected sound change *s > n, via an intermediate stage of ɬ in the following way: Proto-Algonquian *s > ɬ > l > n (Jacques 2013)[22]. Furthermore, a sound change such as */s/ > /ɬ/ implies the simultaneous change of only few phonetic properties, *viz.* [lateral] [continuant]. Phonemes such as /*ts-/ and /*tʂ-/ have merged into a plain voiceless dentalveolar stop /t/, while /*t-/ and /*d-/ have completely disappeared. The aspirated dentalveolar stop /*tʰ-/ has debuccalised into the abutting voiced segment /h/, but this sound change is also observed in Xīnhuì Hécūn, Jiāngmén, Dòumén (Ng Yap) and Nánhǎi (non-Ng Yap, see Péng 1990). According to Zhāng Wèigāng (1943), in the Tàihé dialect of Jiāngxī both *duān* (*t-) and *tòu* initials have apparently become /h/.

Since Yue dialects exhibit only a limited number of morphological processes, in this paper much attention has been given to phonemics, because "phonemic mergers are clearly innovations" (Ringe et al. 2002: 70). If in biology two species are shown to be greatly diverse based on all those features which are controlled by anatomy, such as growth, metabolism, behaviour, etc., then *mutatis mutandis* two languages may be proven to be highly diverse (i.e. separated) on the grounds of those features which are controlled by phonemics, such as aspiration, tonal behaviour, etc.

3.1. Ng Yap innovations (tones)

A description of the tone classes of the Xīnníng dialect was already provided by Don (1883). It seems that very little, if anything, has changed since his times. The tonal behaviour of the various Ng Yap dialects resembles much more that of the Northern varieties of Chinese, than that of other Yue dialects. Like Guǎngfǔ and Gāoyáng, Ng Yap dialects aspirate in the lower rising and in the even tones, and generally confine occlusive initials to the lower rising tones, but unlike other varieties of Yue, the upper even tone has not assumed a falling cadence, and the upper rising tone has not become a very high level tone. Guǎngfǔ has developed a middle tone for words in the entering tone, while in Ng Yap they are in the upper series, where they presumably belonged at an older stage.

Traditional *yáng* tones are higher in pitch than *yīn* tones,[23] a feature rarely observable in Yue, but quite common in Hakka, Gàn, Wú and Xiāng dialects (Yue-Hashimoto 1988, 1991). A major innovation, observable exclusively in Ng Yap, is that the upper even tones have merged with the lower departing tones[24]. This feature is not observed elsewhere (see Table 2 below).

This phoneme should not be confused with the coronal lateral phoneme /ɬ/, which instead is realised with the back of the tongue raised towards the velum.

[22] For different solutions, see Picard 1994.

[23] The two terms are generally considered to be two labels for 'high' and 'low' (Bauer & Benedict 1997: 121). In fact, the two terms are misnomers, because there are dialects, such as Hakka and Mǐn, in which 'low tones' are actually higher in pitch. Thus, the two terms are in fact two impressionistic labels formerly applied to a given quality of the toneme, probably 'height,' but they should not be absolutised.

[24] Zhān & Cheung (1987) observed that the mid-rising tone of the Yǎyáo dialect (Hèshān) is also high-pitched. However, it seems to the present writer that this is, in all probability, a case of changed tone (*biàn yīn* 變音), as it only concerns few words (see also Yue 1991).

Hence, generally speaking, Ng Yap dialects show certain similarities in tones with other Yue dialects (especially with the Yángjiāng and Yángchūn dialects[25]), but also show distinct innovations.

3.2. Ng Yap innovations (rimes)

Concerning rimes, Ng Yap innovations are not numerous. The two most important innovations are as follows (Yue-Hashimoto 1991):

(a) traditional Grade II, III, IV rimes of the *xiè shè* 蟹攝 have merged. They show no vowel length contrast; thus, characters such as 街 'street' and 雞 'chicken,' which in standard Cantonese are respectively /kaːiɭ/ and /kɐiɭ/, in Ng Yap dialects such as Hoishanese are *gai*$^{21–55}$ and *gai*^{33}.

(b) *xiào shè* 效攝 Grade II rimes have merged with Grade I rimes of *liú shè* 流攝. Hence, Ng Yap dialects show no long vowel vs. short vowel contrast in words which are traditionally assigned to these categories.

Nevertheless, there are two things that we need to make clear at the outset: first, these sound changes occur also in the Yue dialects spoken at Téngxiàn, Ēnhè and Shínán; second, although it is stated that sound class X has changed into Y, we cannot be on a firm footing regarding the actual time when this change occurred. Actually, we cannot even be sure that this sound change did really occur. As stated in sections 2.1 and 3, medieval Chinese classes should not be used as a phlogiston which could provide a mechanistic explanation for ways in which sound changes have occurred in all Chinese dialects. In fact, medieval Chinese sound classes represent a North-South *mixtum compositum* of literary pronunciations from different epochs. Many scholars still continue to work within the framework of medieval Chinese sound classes, but it is a mere scholar convention, and medieval Chinese does not reflect a living language. In fact, it is hard to imagine that all those scholars who are mainly concerned with the assessment of medieval Chinese sound classes have never heard that "chaque mot a son histoire."

Since prevocalic glides have been considered traditionally a part of the rime,[26] vocoid approximants should be discussed in this section. If it is true that the four Grades of rime tables indicated the presence of certain glides, then it seems that in standard Cantonese they have disappeared; in most Ng Yap dialects, the two vocoid approximants /j/ and /w/ are clearly vowel-depending, which is to say that they occur respectively only before /ɛ/ and /ɔ/. This means that they are predictable phonetic onglides. In other Yue dialects which exhibit medial glides, these two vocoid approximants are not so predictable (*cf.* Lín Qīnjuān 2008).

3.3. Ng Yap innovations (initials)

As discussed in section 3, the Ng Yap dialects show a wide range of fortitions and lenitions. Nevertheless, most of these peculiarities, such as prenasalised occlusives or the presence of a voiceless alveolar lateral fricative which contrast with standard Cantonese /s/, are widely diffused among other Yue dialects. Another peculiarity which characterises most of Ng Yap dialects is the merger between traditional *kāikǒu* (with no **-w-*) and *hékǒu* (with **-w-*) *jiàn* 見 /*k-/ and *qī* 溪 /*kʰ-/ initials. But, again, this feature is observed also in Cāngwú, Guìpíng, Língxī, Róngxiàn, Xìnyí, Yángjiāng, Yángchūn, Zhōngshān, Zhūhǎi, *etc*.

[25] Yángchūn and Yángjiāng dialects are very similar to each other but they do show differences: for example, the Yángchūn dialect has only the *yáng rù* tone 53, while the dialect of Yángjiāng has both high upper entering (*shàng yáng rù* 54) and low upper entering (*xià yáng rù* 43) tones. See Liú Wěimín 2012: 17.

[26] In historical Chinese phonology, the rime (*yùn* 韻) may include a medial glide (*yùntóu* 韻頭), a nucleus (*yùnfù* 韻腹) and a coda (*yùnwěi* 韻尾).

Table 2. Distribution of this feature within the Yue subfamily. Data are taken, with minor revisions, from Yue-Hashimoto (1991) and Zhān & Cneung (1987, 1988, 1990, 1994, 1998). The grey slots indicate the Ng Yap branch. Regarding the presence of this feature in the Yángjiāng dialect, Yue-Hashimoto (1991: 169) includes it in her table 1, but does not include it in her further discussion of this feature. To the best of this author's knowledge, this feature is absent in Yángjiāng and Yángchūn, as well as in other non-Ng Yap varieties, with the sole exception being Xīnjiè, clearly a case of parallel development.

feature	dialect	presence/absence
merger of the upper even tone (陰平) with the lower departing (陽去)	廣州 Guǎngzhōu	–
	澳門 Macau	–
	增城 Zēngchéng	–
	花縣 Huā xiàn	–
	從化 Cónghuà	–
	信宜 Xìnyí	–
	南海九江 Nánhǎi jiǔjiāng	–
	順德大良 Shùndé dàliáng	–
	高要 Gāoyào	–
	高明城 Gāomíng chéng	–
	化縣 Huà xiàn	–
	蒼梧（滄州）Cāngwú (Cāngzhōu)	–
	玉林 Yùlín	–
	石南 Shínán	–
	橫縣 Héng xiàn	–
	賓陽 Bīnyáng	–
	南寧平話 Nánníng pínghuà	–
	東莞 Dōngguǎn	–
	寶安 Bǎo'ān	–
	新界 Xīnjiè	+
	廉江 Liánjiāng	–
	惠州 Huìzhōu	–
	北海 Běihǎi	–
	欽州 Qīnzhōu	–
	中山 Zhōngshān	–
	珠海 Zhūhǎi	–
	江門 Jiāngmén	–
	新會城 Xīnhuì chéng	+
	新會河村 Xīnhuì hécūn	–
	台山 Táishān	+
	開平 Kāipíng	+
	恩平 Ēnpíng	–
	鶴山 Hèshān	–
	斗門鎮 Dòumén zhèn	+
	陽江 Yángjiāng	+ (?)
	陽春 Yángchūn	–
	羅定（思賀）Luódìng (Sīhè)	–
	桂平江口 Guìpíng jiāngkǒu	–
	博白 Bóbái	–

There is, however, a feature which is exhibited only by Ng Yap dialects such as Hoishan, Kāipíng and Hèshān. Whereas all other dialects have *t-* or other consonantal initials (traditional *duān* initials /*t-/), Ng Yap has zero. Furthermore, if we exclude the dialect of Nánhǎi, and if we are forced to justify and rescue medieval Chinese consonantism, then we are advised to regard the vowel-like abutting segment *h-* of most Ng Yap dialects (including Dòumén, Ēnpíng, Jiāngmén, Xīnhuì) as a result of the following sound change: *t^h- > *h-*. This trend of dropping tenuis while preserving only their suprasegmental feature /*-^{h}/ is found only in Ng Yap varieties.

3.4. Ng Yap innovations (morphology)

One of the most interesting features exhibited by Ng Yap is the absence of plural markers for expressing plurality in personal pronouns.

Table 3. Hoishanese and Cantonese personal pronouns

person	singular			plural		
	Hoishanese		Cantonese	Hoishanese		Cantonese
	romanisation	IPA	jyut pin	romanisation	IPA	jyut pin
1st	ŋgoi (我)	[ŋɔɪ˧]	ngo5	ŋgoi (吾/呆/我)	[ŋɔɪ˧˩]	ngo5 dei6 (我哋)
2nd	ni (你)	[nɪ˧]	nei5	niek (汝/聶/偌)	[nɪɛk˧˩]	nei5 dei6 (你哋)
3rd	kui (佢)	[k^hʊɪ˧]	keoi5	kiek (劇/佉)	[k^hɪɛk˧˩]	keoi5 dei6 (佢哋)

As can be seen from Table 3, plurality is expressed in Hoishanese (as well as in other Ng Yap dialects) by a change in tone. This phenomenon is observable also in the dialect of Yángjiāng, but contrary to Ng Yap, which shows an "anomaly" in the first person plural, the Yángjiāng dialect expresses 'we' with /ŋɔk˧/. Long ago, Antoine Meillet (1925: 27) had already stressed the importance of "les formes anomales," therefore it is in the opinion of the present writer that this 'exception' deserves much more attention than it has received.

3.5. Ng Yap innovations (lexicon)

Lexical analysis is often overlooked in historical linguistics. Although this author agrees on the fact that lexical analysis alone is not sufficient as a criterion for subgrouping, and that the testimony of morphology and phonemics is also required, the dismissal of lexical analysis a priori is certainly exaggerated.

Ng Yap dialects show a different set of interrogative pronouns for 'who' and 'which':

Table 4. Comparison of Ng Yap interrogative pronouns. Data are taken, with some revisions, from Yue-Hashimoto (1991: 175).

pronoun	Ng Yap	Yángjiāng	Zhōngshān	Gāozhōu	Huàxiàn	Téngxiàn
who	sui^{55} 誰	/mɐt˥ sœy˧˩/ 乜誰	/pin˥ sœy˧˩ / pin 誰	/mɐt˥ sœy˧˩/ 乜誰	/mɐt˥ sœy˧˩/ 乜誰	/mɐt˥ sœy˧˩/ 乜誰
which	nai^{21} gɔi^{33}哪一個	/pin˧˩ kɔ:˧/ pin個	/pin˥ kɔ:˧/ pin個	/sɤœ: nit˥/	/sen˥ tsɛ:k˧/ sen一隻	/bin˥ kɔ:˧/ bin個

Hoishanese (and Ng Yap dialects in general) makes use of the morpheme *həu*[22] 毛 for 'head hair,' like Southern Gàn, instead of the neutral lexeme /fa:tɬ/ 髮. The use of the demonstrative pronoun *kɔi*[21] 該 for 'this,' instead of standard Cantonese /ni:ˀl/ 呢 (probably a Taic loanword, *viz.* níi < Proto-Tai *naj^{c}), may suggest that Hoishanese has either replaced the old borrowing or that it has not borrowed the demonstrative pronoun for some unclear reason.

4. Final considerations

It is interesting to note that whichever are the criteria (including mutual intelligibility) utilised to determine which languages are more closely related to one another within the Yue subfamily, Ng Yap dialects appear to be a distinct, *sui generis* branch. Incidentally, no scholar, to the best of this author's knowledge, has ever demonstrated that the Ng Yap dialects are effectively a branch of the Yue family: it is, therefore, left to the sceptical audience to prove otherwise, a fallacious type of argumentation (*onus probandi incumbit ei qui dicet non ei qui negat*). Given that the Yue family has never been classified by means of the standard methodologies and procedures recommended by historical linguistics, one cannot but wonder why the Ng Yap dialects have always been implicitly considered a branch of Yue, in spite of the wide range of distinctions they show[27]. Since proto-Yue has never been reconstructed, and since no scholar, not even McCoy, has ever demonstrated how Ng Yap forms are effectively later, changed versions of earlier proto-Yue forms, we cannot just claim that Ng Yap is a branch which has simply undergone more radical changes than the other varieties of Yue[28].

But languages do not develop in a vacuum: they are socially, culturally and politically connected with their speakers. Thus, the question which now arises is: where do speakers of Hoishanese come from? The answer is unclear and certainly premature. In biology, when naturally selected features become so differentiated that two subsets are unable to reproduce with each other, we are forced to consider the two types to have developed into two separate species. Similarly, in linguistics, when a given language, which is spoken over any significant area, gradually differentiates and ends breaking up into rather distinct varieties, we may encounter regional dialects of that language which, given sufficient time, may become so different from one another that we are forced to regard them as separate languages. Hence, either

[27] During the 23rd International Conference on Yue dialects, many specialists agreed with the present writer in thinking that Ng Yap may not be a variety of Yue, with some of them even claiming that /kʰɵyɻɬ mɻɻ hɐiɬ jy:tɬ jy:ɻɬ/ 'it is not Yue.' Others (few) completely rejected the idea, claiming that it cannot be proved that Ng Yap are not a branch of Yue, a claim which only shows an amazing lack of understanding, for scholars of such calibre, of the scope of the present paper. Those who were broadly sympathetic towards the argumentations of the present manuscript appeared to be, nonetheless, reluctant to the idea that Ng Yap may not be a variety of Yue, although they had to admit that the distinctions and the idiosyncratic features exhibited by Ng Yap dialects are not easily explainable. This writer may be mistaken, but it seems that the refutation of such a conclusion is motivated only by the desire of preserving at all costs the traditional subgrouping, which recognises the existence of Xiāng, Gàn, Mǐn, Wú, Hakka and Yue as the only varieties of Southern Sinitic.

[28] Perhaps, regionalism and provincialism have played a role in this game, reinforced by some lingering adumbrations of the normally involved and generally rather special political unity that has allegedly existed in many parts of the present-day Chinese territory. Moreover, empty concepts, invented in the recent years in the — nonetheless courageous and remarkable — attempt of replacing the English mistranslation of the Chinese word *fāngyán* 方言, such as 'regionalect' (*cf.* DeFrancis 1984: 57) and 'topolect' (*cf.* Mair 1991: 7, 2008) have probably added further fuel to the fire, with the result that, paradoxically, this undemonstrated classification has hardened into an orthodoxy that none have dared to challenge.

Ng Yap once was really a branch of proto-Yue whose evolution, nonetheless, involved linguistic changes so fundamental that now they should be considered to be different groups, or is a branch of another subfamily which was radically transformed under a Yue substratum when its speakers moved to the south-western coast of the Yue speaking area. Of course, another possible answer is that all the doubts expressed in this paper are circumstantial and not worth considering, and that Ng Yap is effectively and without any hint of doubt a branch of Yue. Solution of this problem goes vastly beyond the scope of the present paper, but if this work is allowed to dwell a little longer on this topic, then one might also hypothesise that the parent language of the various Ng Yap dialects was the language historically spoken by Song soldiers who were stationed in today's Guǎngxī, Fújiàn and Guǎngdōng provinces during the Southern Song epoch (1127–1279), possibly as a result of the loss of the capital of Hángzhōu at the hands of Mongol invaders. The presence of words such as *ɔn*[33] *ŋin*[22] 安人 for 'husband's mother,' which is the courtesy form used for officials' family during the Song dynasty,[29] is very suggestive, though still hardly sufficient to prove this scenario.

Be it as it may, Ng Yap vocalism resembles Hakka, its tonal behaviour is closer to the Northern varieties of Mandarin than to that of other Yue dialects, and it shows predictable phonetic onglides, unlike any other variety of Yue. Its consonantism also shows two unique features, namely the loss of tenuis and voiced dentalveolar plosives and the debuccalisation of an aspirated dentalveolar stop. The use of personal and interrogative pronouns also is much closer to Northern varieties of Mandarin than to any other Yue dialect.

The humble aim of this paper is to claim that Ng Yap shows a wide range of distinct features which *may* warrant its separation. Subgrouping or the placing of a given language within a family is inevitably a matter of weighing criteria on an arbitrary basis: one has to choose which features of a given language are the most important, and of course the perspective that one adopts inescapably changes the weighing that one gives. The present paper has chosen to give more emphasis to specific features of tonal behaviour, as well as to the morphological process involving the pluralisation of personal pronouns. Other phonemic features, such as the treatment of certain initial consonants and the presence of prevocalic glides, and a few lexical features, such as the use of interrogative pronouns, have also received special attention.

In concluding, although further effort is needed to strengthen the conclusions drawn in this paper,[30] it is in the opinion of the present writer that any other conclusion will require a whole lot of special pleading.

References

Ang, Uijin (洪惟仁). 1996. *Huìyīn miàowù yǔ gǔdài Quánzhōu yīn* (彙音妙悟與古代泉州音 [The Lūi-im Biāu-gō͘ and the sound system of the ancient Quánzhōu language]). Guólì zhōngyāng túshūguǎn Táiwān fēnguǎn (國立中央圖書館臺灣分館).

Atkinson, Quentin, Russell Gray. 2006. In: Peter Forster, Colin Renfrew (eds.). *How old is the Indo-European language family? Illumination or more moths to the flame? Phylogenetic methods and the prehistory of languages*: 91–110. Cambridge: MacDonald Institute Press.

Ball, James Dyer. 1890. The Tung-Kwun (東莞) Dialect. *The China Review* 18(4): 284–299.

[29] See the *Dà Sòng xuānhé yíshì* 大宋宣和遺事: http://open-lit.com/html/lit/562/19250.html.

[30] A good way might be to write down on a spreadsheet all the peculiar features exhibited by the various Yue languages, *e.g.* aspiration in the upper range of tones 1, merger of upper even with upper departing tones 2, etc., and to put it into a computer software, such as PAUP, which uses an algorithm to create a phylogenetic tree. This method is widely diffused in many areas of science, such as palaeontology (see, for instance, Brusatte et al. 2010).

Ball, James Dyer. 1907. *Cantonese Made Easy: A book of simple sentences in the Cantonese dialect, with free and literal translations, and directions for the rendering of English grammatical forms in Chinese*. Singapore: Kelly & Walsh.

Bauer, Robert S., Paul K. Benedict. 1997. *Modern Cantonese phonology*. Vol. 102. Berlin / New York: Walter de Gruyter.

Baxter, William H. 2006. Mandarin dialect phylogeny. *Cahiers de Linguistique — Asie Orientale* 35 (1): 71–114.

Bickel, Balthasar. 2014. Linguistic diversity and universals. In: N. J. Enfield, Paul Kockelman, Jack Sidnell (eds.). *The Cambridge Handbook of Linguistic Anthropology*: 101–124. Cambridge University Press.

Brusatte, Stephen L., Michael J. Benton, Julia B. Desojo, Max C. Langer. 2010. The higher-level phylogeny of Archosauria (Tetrapoda: Diapsida). *Journal of Systematic Palaeontology* 8(1): 3–47.

Campbell, Lyle, William J. Poser. 2008. *Language classification: history and method*. Cambridge University Press.

Chao, Yuen-Ren. 1928. *Studies in the Modern Wu Dialects*. Tsinghua College Research Institute Monograph, 4. Beijing.

Chao, Yuen-Ren. 1943. Languages and dialects in China. *The Geographical Journal* 102(2): 63–66.

Chao, Yuen-Ren. 1951. Taishan Yuliao (台山語料 [Corpus of Toishanese]). *Bulletin of the Institute of Historical Philology* (*Academia Sinica*) 23: 25–76.

Chén, Ruìqīng 沈瑞清. 2018. *Huángkēnghuà yǔ zǎoqí Mǐn běi yǔ* (黃坑話與早期閩北語 [The Huángkēng dialect and early Northern Mǐn]). Talk given at the The 2nd Li Fanggui Young Scholars Seminar (Institute of Languages, Academia Sinica).

Cheng, Chin-Chuan. 1982. A quantification of Chinese dialect affinity. *Studies in the Linguistic Sciences* 12(1): 29–47.

Cheng, Chin-Chuan. 1986. *Quantifying Affinity Among Chinese Dialects*. Language Learning Laboratory Technical Report LLL-T -10–86. Urbana, Illinois: University of Illinois.

Cheng, Chin-Chuan. 1987. A quantification of Chinese dialect affinity. In: Chinese Language Society of Hong Kong (ed.), *Wang Li Memorial Volumes English Volume*: 73–100. Hong Kong: Joint Publishing.

Cheng, Chin-Chuan. 1988. Hànyǔ fāngyán qīnshǔ guānxì de jìliàng yánjiū (漢語方言親屬關係的計量研究 [Quantitative studies of Chinese dialect affinity]). *Zhōngguó yǔwén* (中國語文 [Studies of the Chinese Language]) 203: 87–102.

Cheng, Chin-Chuan. 1991. Quantifying affinity among Chinese dialects. *Journal of Chinese Linguistics Monograph Series* 3: 76–110.

Cheng, Chin-Chuan. 1992. Syllable-based dialect classification and mutual intelligibility. In: *Chinese Languages and Linguistics (I): Chinese Dialects. Symposium Series of the Institute of History and Philology, Academia Sinica 2*: 145–177.

Cheng, Chin-Chuan. 1994. Hànyǔ fāngyán gōutōng dù de jìsuàn (漢語方言溝通度的計算 [Calculation of Chinese dialectal mutual intelligibility]). *Zhōngguó yǔwén* (中國語文 [Studies of the Chinese Language]) 238: 35–43.

Cheng, Chin-Chuan. 1996. Quantifying dialect mutual intelligibility. In: C.-T. James, Huang Yen-hui, Audrey Li (eds.). *New horizons in Chinese linguistics*: 269–292. Dordrecht: Springer.

Cheng, Teresa M. 1973. The Phonology of Taishan. *Journal of Chinese Linguistics* 1(2): 256–322.

D'Andrea, John A., Timothy Light. 1981. Toishan Glossary Project, an Interim Report. *Journal of Chinese Linguistics* 9(1): 130–137.

Defense Language Institute. 1964. *Chinese-Cantonese (Toishan) Basic Course*. Washington, DC: Defense Language Institute.

DeFrancis, John. 1984. *The Chinese language: Fact and fantasy*. Honolulu: University of Hawaii Press.

Dèng, Jūn (鄧鈞). 2006. *Táishān fāngyīn zìdiǎn* (台山方音字典 [Phonetic dictionary of Hoishanese]). Chángshā: Húnán jiàoyù chūbǎnshè (長沙:湖南教育出版社).

Dèng, Jūn 鄧鈞 (ed.). 2000. *Kāipíng fāngyán (fù yīndài)* (開平方言(附音帶) [The Kaiping dialect (with an appendix)]). Chángshā: Húnán diànzǐ yīnxiàng chūbǎn shè (長沙:湖南電子音像出版社).

Don, Alexander. 1883. The Llin-nen variation of Cantonese. *China Review* 11: 236–247.

Don, Alexander. 1884. The Llin-nen variation of Cantonese. *China Review* 12: 474–481.

Douglas, Carstairs. 1873. *Chinese-English dictionary of the vernacular or spoken language of Amoy, With the principal variations of the Chang-chew and Chin-chew dialects*. Trübner [Suppl.:] Commercial Press in Shanghai.

Dunn, Michael, Robert Foley, Stephen Levinson, Ger Reesink, Angela Terrill. 2007. Statistical reasoning in the evaluation of typological diversity in Island Melanesia. *Oceanic Linguistics* 46(2): 388–403.

Eiichi, Chishima (千島英一). 2007. *Honkon Kanton-go no kenkyū: Goi to gengo bunka o chūshin ni* (香港広東語の研究: 語彙と言語文化を中心に [A study of Hong Kong Cantonese from a viewpoint of vocabulary, language and culture]). Ph.D. thesis: Kanazawa Daigaku daigakuin shakai kankyō kagaku kenkyū-ka (金沢大学大学院社会環境科学研究科 [Graduate School of Social and Environmental Sciences, Kanazawa University]).

Eska, Joseph F., Donald A. Ringe. 2004. Recent work in computational linguistic phylogeny. *Language* 80(3): 569–582.

Evans, Steven N., Don Ringe, Tandy Warnow. 2006. Inference of divergence times as a statistical inverse problem. In: Peter Forster, Colin Renfrew (eds.). *Phylogenetic methods and the prehistory of languages*: 119–130. Cambridge: McDonald Institute for Archaeological Research.

Faith, Daniel P., Catherine A. Lozupone, David Nipperess, Rob Knight. 2009. The cladistic basis for the phylogenetic diversity (PD) measure links evolutionary features to environmental gradients and supports broad applications of microbial ecology's "phylogenetic beta diversity" framework. *International Journal of Molecular Sciences* 10(11): 4723–4741.

Farris, James S., Arnold G. Kluge, Michael J. Eckardt. 1970. A numerical approach to phylogenetic systematics. *Systematic Zoology* 19(2): 172–189.

Féng, Guóqiáng (馮國強) (ed.). 2012. *Sháoguān shìqū Yuèyǔ yǔyīn biànyì yánjiū* (韶關市區粵語語音變異研究 [Research on Cantonese Phonetic Variation in Shaoguan City]). Wànjuǎnlóu chūbǎnshè (萬卷樓出版社).

Forrest, Robert Andrew Dermod. 1948 [1965]. *The Chinese Language*. London: Faber & Faber.

Gān, Yú'ēn (甘於恩), Zēng Jiànshēng (曾建生). 2010. Guǎngdōng dìlǐ yǔyánxué yánjiū zhī ruògān sīkǎo (廣東地理語言學研究之若干思考 [Some Thoughts on the Study of Geographical Linguistics in Guangdong]). *Jìnán xuébào (zhéxué shèhuì kēxué bǎn)* (暨南學報哲學社會科學　版) 3: 20–27.

Gān, Yú'ēn (甘於恩). 2003. Sìyì fāngyán de xíngtài biàndiào (四邑方言的形態變調 [The morphological changes of the Sze Yap dialects]). In: Zhan Bohui (ed.). *Dì bā jiè guójì Yuè fāngyán yántǎohuì lùnwénjí* (第八屆國際粵方言研討會論文集 [Proceedings of the 8th International Cantonese Dialect Symposium]): 227–250. Beijing: Zhōngguó shèhuì kēxué chūbǎnshè (中國社會科學出版社).

Gān, Yú'ēn (甘於恩). 2008. Guǎngdōng liǎng Yáng Yuèyǔ yǔyīn tèdiǎn gàishuō (廣東兩陽粵語語音特點概說 [General description of the main charachteristics of the Yangchun and Yangjiang dialects]). *Guìlín shīfàn gāoděng zhuānkēxuéxiào xuébào* (桂林師範高等專科學校學報) 22(2): 1–6.

Gān, Yú'ēn (甘於恩). 2010. *Guǎngdōng Sìyì fāngyán yǔfǎ yánjiū* (廣東四邑方言語法研究 [Research on the Grammar of Guangdong Siyi Dialect]). Jìnán dàxué chūbǎnshè (暨南大學出版社).

Gray, Russell D., Quentin D. Atkinson. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426(6965): 435–439.

Hashimoto, Mantaro (橋本萬太郎). 1978. *Gengo ruikei-chiri ron* (言語類型地理論 [Language type geography theory]). Tokyo: Kōbudo (弘文堂).

Hashimoto, Mantaro (橋本萬太郎). 2010. *The Hakka dialect: A linguistic study of its phonology, syntax and lexicon*. Vol. 5. Cambridge University Press.

Hé Wénhuì (何文匯). 2009. *Yuèyīn píngzè rùmén-Yuèyǔ zhèngyīn shìlì (hé dìng běn)* (粵音平仄入門-粵語正音示例 [Introduction to Cantonese Tones — Cantonese Correct Sounds]). Míng chuāng (明窗).

Hennig, Willi. 1950. *Grundzuge einer Theorie der phylogenetischen Systematik*. Berlin: Deutscher Zentralverlag. Translated into English in 1966 as *Phylogenetic Systematics*, Annual Review of Entomology 10: 97–116.

Him, Kam Tak. 1980. Semantic-Tonal Processes in Cantonese, Taishanese, Bobai and Siamese. *Journal of Chinese Linguistics* 8(2): 205–240.

Ho, Dah-An (何大安). 1988. Zhuó shǎng guī qù yǔ xiàndài fāngyán (濁上歸去與現代方言 [Ascription of Voiced Rising to Departing Tone and Modem Dialects]). *Zhōngyāng yánjiūyuàn lìshǐ yǔyán yánjiūsuǒ jìkān* (中央研究院歷史語言研究所季刊) 59: 115–140.

Hóu, Xìngquán (侯興泉). 2016. Yuèyǔ gǔ quánzhuó shēngmǔ jīndú zhuóyīn de lèixíng jí chéngyīn — jiānlùn gǔ quánzhuó shēngmǔ jīndú zài fāngyán fēnqū zhōng de zuòyòng (粵語古全濁聲母今讀濁音的類型及成因 — 兼論古全濁聲母今讀在方言分區中的作用 [The Types and Causes of the Voiced Sounds of the Ancient Cantonese Voiced Initials — On the Role of the Ancient Voiced Initials in the Dialect Division]). *Yǔyán jì yǔyánxué* (語言暨語言學) 17(5): 661–678.

Hsie, Hsin-I (謝信一). A new method of dialect subgrouping. *Journal of Chinese Linguistics* 1(1): 64–89.

Hú, Hǎiqióng (胡海瓊). 2012. Cóng chūtǔ wénxiàn kàn shànggǔ Chǔyǔ de xìngzhì (從出土文獻看上古楚語的性質 [On the nature of ancient Chu language from the unearthed literature]). *Xuéxí Yuèkān* (學習月刊) 16: 10–16.

Huáng, Jiànyún (黃劍雲) (ed.), Yè Guóquán (葉國泉) (rev.). 1990. *Táishān fāngyán* (台山方言 [The Taishan Dialect]). Guǎngzhōu: Zhōngshān dàxué chūbǎnshè (中山大學出版社).

Jacques, Guillaume. 2013. The sound change *s- > n- in Arapaho. *Folia Linguistica Historica* 34(1): 43–58.

Jacques, Guillaume, Johann-Mattis List. 2018. Save the trees: Why we need tree models in linguistic reconstruction (and when we should apply them). *Journal of Historical Linguistics* 8(2): 128–166.

Karlgren, Bernhard. 1915–1926. *Études sur la phonologie chinoise*. Archives d'Etudes Orientales.

Krishnamurti, Bhadriraju, Lincoln Moses, Douglas G. Danforth. 1983. Unchanged cognates as a criterion in linguistic subgrouping. *Language* 59(3): 541–568.

Kwok, Bit Chee (郭必之). 2006. Yīduàn wéiguān de Yuèyǔ shǐ: Zhūjiāng sānjiǎo jiāng bā zhǒng yuèyǔ fāngyán de fēnqún (一段微觀的粵語史：珠江三角江八種粵語方言的分群 [A microscopic Cantonese history: the grouping of eight Cantonese dialects in the Pearl River Delta]). In: Ho Dah-an (ed.) *Shān gāo shuǐ cháng: Dīng Bāngxīn xiānshēng qī zhì shòuqìng lùnwénjí* (山高水長: 丁邦新先生七秩壽慶論文集 [High as mountains and long as rivers: Festschrift in honor of Ting Pang-hsin for his seventeenth birthday]): 701–739. Táiběi: Zhōngyāng yánjiuyuàn yǔyánxué yánjiū suǒ (中央研究院語言學研究所).

Kwok, Bit Chee, Andy C. Chin, Benjamin K. Tsou. 2016. Grammatical diversity across the Yue dialects. *Journal of Chinese Linguistics* 44(1): 109–152.

Lau, Chun-fat (劉鎮發), Liáng Huìmǐn (梁慧敏). 2011. Zhūjiāng sānjiǎozhōu fāngyán yīn rùshēng shòu Guǎngzhōuhuà yǐngxiǎng de yǎnbiàn qíngkuàng (珠江三角洲方言陰入聲受廣州話影響的演變情況 [The Cantonese influence over the evolution of the lower entering tone in the Pearl River Delta dialects]). *Yǔyán yánjiū* (語言研究 [Studies in Language and Linguistics]) 4: 92–97.

Lau, Chun-fat (劉鎮發). 2002. Hànyǔ fāngyán de fēnlèi biāozhǔn yǔ "kèjiā huà" zài hànyǔ fāngyán fēnlèi shàng de wèntí (漢語方言的分類標準與"客家話"在漢語方言分類上的問題 [Criteria for Chinese dialect classification and the problem of the position of the Hakka dialect in Chinese dialect grouping]). *Journal of Chinese Linguistics* 30(1): 82–96.

Lau, Chun-fat (劉鎮發). 2007. Xiàndài fāngyán de [j] zhuócāhuà — fùlùn Zhōnggǔ xiémǔ de xíngchéng (現代方言的[j]濁擦化—附論中古邪母的形成 [The voicing and fricativisation of [j] in the modern dialects—and the formation of the *xié* initial in Middle Chinese]). *Zhōngguó yǔwén* (中國語文) 317(2): 169–174.

Lee, Gina. 1987. A Study of Toishan F0. *Ohio State University Working Papers in Linguistics* 36: 16–30.

Lǐ, Dōngxiāng (李冬香). 2010. Yuèběi tǔhuà xiǎochēng biànyīn de biànyì yánjiū (粵北土話小稱變音的變異研究 [Research on Northern Cantonese changed tones]). *Jìnán xuébào (zhéxué shèhuì kēxué bǎn)* (暨南學報 哲學社會科學版) 32(1): 111–116.

Li, Fang-Kuei. 1937. *Languages and Dialects of China*. The Chinese Yearbook. Shanghai: Commercial Press (Reprinted in Journal of Chinese Linguistics, 1:1–13).

Lǐ, Jǐnfāng (李錦芳). 2002. *Dòng Tái Yǔyán yǔ Wénhuà* (侗台語言與文化 [Tai-Kadai Language and Culture]). Beijing: Mínzú chūbǎnshè (民族出版社).

Lǐ, Jìngzhōng (李敬忠). 1990. Yuèyǔ shì Hànyǔzú qún zhōng de dúlì yǔyán (粵語是漢語族群中的獨立語言 [Cantonese as an independent language in the family of Chinese languages].) In: Zhan Bohui (ed.). *Dì èr jiè guójì Yuè fāngyán yántǎo huì lùnwén jí* (第二屆國際粵方言研討會論文集 [Proceedings of the 2nd International Cantonese Dialect Symposium]): 22–29. Guǎngzhōu: Jìnán dàxué chūbǎnshè (暨南大學出版社).

Lǐ, Lìlín (李立林). 2010. *Dōngguǎn Yuèyǔ yǔyīn yánjiū* (東莞粵語語音研究 [Dongguan Cantonese Speech Research]). Ph.D. Thesis. Guǎngzhōu: Jìnán University.

Light, Timothy. 1986. Toishan affixal aspects. In: John McCoy, Timothy Light (eds.). *Contributions to Sino-Tibetan studies*: 415–425. Leiden: Brill.

Lín, Bǎisōng (林柏松). 1990. Shùndéhuà zhōng de "biàn yīn" (順德話中的"變音" [Changed tones in Shunde dialect].). In: Zhan Bohui (ed.). *Dì èr jiè guójì Yuè fāngyán yántǎo huì lùnwén jí* (第二屆國際粵方言研討會論文集 [Proceedings of the 2nd International Cantonese Dialect Symposium]): 86–88. Guǎngzhōu: Jìnán dàxué chūbǎnshè (暨南大學出版社).

Lín, Qīnjuān (林欽娟). 2008. Qīnzhōuhuà tóngyīnzìhuì (欽州話同音字彙 [The homophony of the Qinzhou dialect]). *Guìlín shīfàn gāoděng zhuānkē xuéxiào xuébào* (桂林師範高等專科學校學報) 22(1): 1–7.

Liu, Fu (Liu Bannong). 1925. *Étude expérimentale sur les tons du chinois*. Paris: Société d'Éditions "Les Belles Lettres."

Liú, Wěimín (劉偉民). 2012. Yángchūnhuà yǔ Yángjiānghuà de yǔyīn chāyì (陽春話與陽江話的語音差異). *Xiàndài yǔwén, yǔyán yánjiū* (現代語文, 語言研究) 6: 16–18.

Liú, Xīnzhōng (劉新中), Yáng Wèi (楊蔚), Cuī Shūhuì (崔淑慧). 2003. Guānyú hànyǔ fāngyán fēnqū de yǔyán tiáojiàn (關於漢語方言分區的語言條件 [Linguistic criteria for the classification of Chinese dialects]). *Yǔyán yánjiū* (語言研究) 3: 113–118.

Luó, Yánfā (羅言發). 2013. *Àomén Yuèyǔ yīnxì de lìshǐ biànqiān jí qí chéngyīn* (澳門粵語音系的歷史變遷及其成因 [The Historical Changes of the Macao Cantonese Phonetic System and Its Causes]). Ph.D. Thesis: Běijīng University.

Maddieson, Ian, Sandra Ferrari Disner. 1984. *Patterns of sounds*. Cambridge University Press.

Mài, Yún (麥耘). 2010. Yuèyǔ de Xíngchéng, fàzhǎn yǔ Yuèyǔ hé Pínghuà de Guānxì (粵語的形成、發展與粵語和平話的關係 [The formation and development of Yue, and the relationship between Yue and Pinghua]). In: Pān Wùyún (潘悟云), Shěn Zhōngwěi (沈鍾偉) (eds.). *Yánjiū zhī lè — qìngzhú Wáng Shìyuán xiānshēng qīshíwǔ shòuchén xuéshù lùnwénjí* (研究之樂慶祝王士元先生七十五壽辰學術論文集 [The Joy of Research II—A Festschrift in Honor of Professor William S-Y. Wang on His Seventy-fifth Birthday]): 227–243. Shànghǎi: Shànghǎi jiàoyù chūbǎnshè (上海教育出版社).

Mair, Victor H. 1991. *What is a Chinese "dialect/topolect"?: Reflections on Some Key Sino-English Linguistic Terms*. Philadelphia: Department of Oriental Studies, University of Pennsylvania.

Máo, Yúháng (茅于杭), Zhāng Guózhēn (張國貞). 2001. Guǎngdōnghuà wénběn yǔyīn zhuǎnhuàn xìtǒng de shèjì yǔ shíxiàn (廣東話文本語音轉換系統的設計與實現 [Design and Implementation of Cantonese Text-to-speech System]). In: Huáng Chāngnìng (ed.). *Zìrán yǔyán lǐjiě yǔ jīqì fānyì——quánguó dì liù jiè jìsuàn yǔyánxué liánhé xuéshù huìyì lùnwénjí* (自然語言理解與機器翻譯 — 全國第六屆計算語言學聯合學術會議論文集 [Natural Language Understanding and Machine Translation — Proceedings of the 6th National Conference on Computational Linguistics]): 443–447. Qīnghuá dàxué chūbǎnshè (清华大学出版社).

Matisoff, James A. 2000. On the uselessness of glottochronology for the subgrouping of Tibeto-Burman. In: Colin Renfrew, April McMahon, Larry Trask (eds.). *Time depth in historical linguistics*: 333–372. Cambridge: The McDonald Institute for Archaeological Research.

McCoy, John. 1966. *Szeyap Data for a First Approximation of Proto-Cantonese*. Ph.D. Thesis: Cornell University.

Marcet-Houben, Marina, Toni Gabaldón. 2011. TreeKO: a duplication-aware algorithm for the comparison of phylogenetic trees. *Nucleic acids research* 39(10): e66.

Minett, James W., William S.-Y. Wang. 2003. On detecting borrowing: distance-based and character-based approaches. *Diachronica* 20(2): 289–330.

Mueller, Laurence D., Francisco J. Ayala. 1982. Estimation and interpretation of genetic distance in empirical studies. *Genetics Research* 40(2): 127–137.

Nakhleh, Luay, Don Ringe, Tandy Warnow. 2005. Perfect phylogenetic networks: A new methodology for reconstructing the evolutionary history of natural languages. *Language* 81(2): 382–420.

Nicholls, Geoff K., Russell D. Gray. 2008. Dated ancestral trees from binary trait data and their application to the diversification of languages. *Journal of the Royal Statistical Society: Series B* (Statistical Methodology) 70(3): 545–566.

Nichols, Johanna, Tandy Warnow. 2008. Tutorial on computational linguistic phylogeny. *Language and Linguistics Compass* 2(5): 760–820.

Norman, Jerry. 1988. *Chinese*. Cambridge University Press.

Ogawa, Naoyoshi (小川尚義). 1931–1932. *Jittai dai kotobaten* (日臺大辭典 [Great Japanese-Taiwanese Dictionary]). Mǐnnányǔ jīngdiǎn císhū huìbiān (閩南語經典辭書彙編 [Compilation of Southern Min Classical dictionaries]), Vol. 5. Táiběi: Táiwān zǒngdūfǔ mínzhèngbù zǒngwùjú xuéwùkè fāxíng (台灣總督府民政部總務局學務課發行).

Peng, Hsin-yi. 2014. The Yue Sub-Dialect Yayao: The Process of and Phonetic Motivation for the Loss of Its Bilabial Obstruent Initials. *Hanxue yanjiu* 32(3): 227–255.

Péng, Xiǎochuān (彭小川). 1990. Guǎngdōng Nánhǎi (shā tóu) fāngyán yīn xì (廣東南海 (沙頭) 方言音系 [The sound system of the Nánhǎi dialect (Shātóu)]). *Fāngyán* (方言) 1(1990): 23–33.

Picard, Marc. 1994. *Principles and methods in historical phonology: From Proto-Algonkian to Arapaho*. Montreal & Kingston: McGill-Queen's University Press.

Pike, Kenneth Lee. 1943. *Phonetics: a critical analysis of phonetic theory and a technique for the practical description of sounds*. Ann Arbor: University of Michigan Press.

Qū, Dàjūn (屈大均). 2006. *Guǎngdōng xīnyǔ* (廣東新語). Běijīng: Zhōnghuá shūjú (中華書局).

Rasmussen, Jens Elmegård. 1991. Trubetzkoy's Thoughts about Indo-European: A myth come into fashion. In: H. A. Sigurdsson (ed.). *Papers From the Twelfth Scandinavian Conference of Linguistics*: *Reykjavík, June 14–16, 1990*: 345–356. Linguistic Institute, University of Iceland.

Ringe, Don, Tandy Warnow, Ann Taylor. 2002. Indo‐European and computational cladistics. *Transactions of the philological society* 100(1): 59–129.

Robinson, David F., Leslie R. Foulds. 1981. Comparison of phylogenetic trees. *Mathematical biosciences* 53(1–2): 131–147.

Róng, Huìhuá (容慧華). 2009. Hétánghuà yǔ Sìyìhuà de bǐjiào yánjiū (荷塘話與四邑話的比較研究 [A comparative study of the Hetang dialect and the Sze Yap dialects]). *Xiàndài yǔwén* (現代語文) 33: 103–105.

Róng, Huìhuá (容慧華). 2010. Hétánghuà xì shǔ wèntí chūtàn (荷塘話係屬問題初探 [A preliminary study on the problem regarding the affiliation of the Hetang dialect].) *Jiāmùsī jiàoyù xuéyuàn xuébào* (佳木斯教育學院學報 [Journal of Jiamusi Education College]) 4: 265–266.

Sagart, Laurent. 1993. Les dialectes gan. *Revue bibliographique de sinologie* 11: 261–262.

Sagart, Laurent. 1998. On Distinguishing Hakka and non-Hakka Dialects (關於客家方言的區別性特徵). *Journal of Chinese linguistics* 26(2): 281–302.

Sagart, Laurent. 2002. Gan, Hakka and the formation of Chinese dialects. In: Dah-an Ho (ed.). *Dialect Variations in Chinese*: 129–154. Academia Sinica.

Shào, Huìjūn (邵慧君), Gān, Yú'ēn (甘於恩). 1999. Guǎngdōng sìyì fāngyán de yǔyīn tèdiǎn (廣東四邑方言的語音特點 [The phonetic features of the Sze Yap dialects]). *Fāngyán* (方言) 3(1999): 128–135.

Shào, Huìjūn (邵慧君). 2010. Lùn Yuè fāngyán i, u jiè yīnyùnmǔ (論粵方言 i, u 介音韻母 [On the medial i, u in modern Cantonese]). *Jìnán xuébào (zhéxué shèhuì kē xué bǎn)* (暨南學報哲學社　會科學版) 6: 116–122.

Simmons, Richard VanNess. 1999. *Chinese dialect classification: a comparative approach to Harngjou, Old Jintarn, and Common Northern Wu*. Amsterdam/Philadelphia: John Benjamins.

de Sousa, Hilário. 2015. Language contact in Nanning: Nanning Pinghua and Nanning Cantonese. In: Hilary Chappell (ed.). *Diversity in Sinitic languages*: 157–189. Oxford University Press.

Stöver, Ben C., Kai F. Müller. 2010. TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics* 11(1): 7.

Takekoshi, Minako (竹越美奈子). 2017. Kindai etsugo gū Tsutomu Osamu ittō ji no hensen (近代粵語遇效攝一等字の変遷 [Changes of Grade I xiao rimes words in Modern Cantonese]). *Tōhō-gaku-shi* (東邦学誌) 46(2): 1–25.

Tan, Yutian. 2011. The origin and nature of high rising diminutive changed tone in Siyi dialect. In: Zhuo Jing-Schmidt (ed.). *Proceedings of the 23rd North American Conference on Chinese Linguistics* (NACCL-23): 190–207. Eugene: University of Oregon.

Tang, Chaoju, V. J. J. P. van Heuven. 2007. Mutual intelligibility and similarity of Chinese dialects: Predicting judgments from objective measures. *Linguistics in the Netherlands* 24(1): 223–234.

Tang, Chaoju, V. J. J. P. van Heuven. 2008. Mutual intelligibility of Chinese dialects tested functionally. *Linguistics in the Netherlands* 25: 145–156.

Tang, Chaoju, V. J. J. P. van Heuven. 2009. Mutual intelligibility of Chinese dialects experimentally tested. *Lingua* 119: 709–732.

Tang, Chaoju, V. J. J. P. van Heuven. 2015. Predicting mutual intelligibility of Chinese dialects from multiple objective linguistic distance measures. *Linguistics in the Netherlands* 53(2): 285–312.

Tang, Chaoju. 2017. Dialects of Chinese. In: Charles Boberg, John Nerbonne, Dominic Watt (eds.). *The Handbook of Dialectology*: 547–558. Wiley-Blackwell Publishers.

Tāng, Yíngyíng (湯瑩瑩). 2011. *Shùndé dàliáng fāngyán de yǔyīn yánjiū* (順德大良方言的語音研究 [Speech Research of Shunde Daliang Dialect]). Master's Thesis, South China Normal University.

Ting, Pang-hsin (丁邦新). 1982. Hànyǔ fāngyán qūfēn de tiáojiàn (漢語方言區分的條件 [Phonological features for classification of the Chinese dialects]). *Tsing Hua Journal of Chinese Studies* 14((1–2): 257–273.

Ting, Pang-hsin (丁邦新), Chang Song-hing (張雙慶) (eds.). 2002. *Mǐnyǔ yánjiū jí qí yǔ zhōubiān fāngyán de guānxì* (閩語研究及其與周邊方言的關係 [The study of Mǐn and their relationship with surrounding dialects]). Hong Kong: Chinese University of Hong Kong Press.

Tong, Choi Lan (湯翠蘭). 1997. *Sìyì fāngyán shēngyùn chūtàn* (四邑方言聲韻初探 [A preliminary study of the phonology of Sze Yap dialects]). PhD diss., National Taiwan University Department of Chinese Literature.

Tong, Choi Lan (湯翠蘭). 2004. *Guǎngdōng Xīnhuì Sīqiánhuà de yīnyùn yánjiū* (廣東新會司前話的音韻研究 [The phonology of the Sīqián dialect of Xīnhuì (Guǎngdōng)]). PhD diss., Jìnán University.

Trask, Robert Lawrence. 2007 [1999]. *Language and linguistics: The key concepts*. 2nd edition. London/New York: Routledge.

Wáng, Lì (王力), Qián Sōngshēng (錢淞生). 1930. Táishān fāngyīn (台山方音 [Sounds of Taishanese]). *Lǐngnán xuébào* (嶺南學報) 10(2): 67–104.

Wáng, Lì (王力), Qián Sōngshēng (錢淞生). 1950. Zhūjiāng sānjiǎozhōu fāng yīn zǒng lùn (珠江三角洲方音總論 [General description of the Pearl River Delta dialects]). *Lǐngnán xuébào* (嶺南學報) 10(2): 57–66.

Wáng, Lì (王力). 1936. *Zhōngguó yīnyùnxué* (中國音韻學 [Phonology of Chinese]). Běijīng: Shāngwù yìnshūguǎn (商務印書館).

Wáng, Màolín (王茂林), Liú Xīnzhōng (劉新中). 2010. Liǎng Yáng Yuèyǔ Yángjiānghuà jǐ gè fǔyīn de yǔyīnxué yánjiū (兩陽粵語陽江話幾個輔音的語音學研究 [The Phonetic Study of Several Consonants in Yangjiang Dialect of Liangyang Cantonese]). In: *Dì jiǔ jiè Zhōngguó yǔyīnxué xuéshù huìyì lùnwénjí* (第九屆中國語音學學術會議論文集 [Proceedings of the 9th Chinese Conference on Phonetics]). Tiānjīn.

Wang, William S.-Y., Shěn Zhōngwěi (沈鐘偉). 1992. Fāngyán guānxi de jìliàng biǎoshù (方言關係的計量表述 [Numerical representations of dialectal relationships]). *Zhōngguó yǔwén* (中國語文) 2: 81–92.

Wang, William S.‐Y., James W. Minett. 2005. Vertical and horizontal transmission in language evolution 1. *Transactions of the Philological Society* 103(2): 121–146.

Wang, William S.-Y. 1987. Representing languages relationships. In: Henry M. Hoenigswald, Linda F. Wiener (eds.). *Biological Metaphor and Cladistic Classification. An Interdisciplinary Perspective*: 243–256. Philadelphia: University of Pennsylvania Press.

Wang, William S.-Y. (ed.). 1991. *Languages and dialects of China (Journal of Chinese Linguistics Monograph Series No. 3)*. Berkeley: Project on Linguistic Analysis.

Wang, William S.-Y. 1994. Glottochronology, lexicostatistics, and other numerical methods. In: R. E. Asher, J. M. Y. Simpson (eds.). *The Encyclopedia of Language and Linguistics Vol. 3*: 1445–1450. Oxford: Pergamon Press.

Wang, William S.-Y. 1996. Linguistic diversity and language relationships. In: C.-T. James, Huang Yen-hui, Audrey Li (eds.). *New horizons in Chinese linguistics*: 235–267. Dordrecht: Springer.

Wáng, Yùdé. 1960. Chūgoku go dai hōgen no bunretsu nendai no gakuteki shitan (中国五大方言の分裂年代の学的試探 [The lexicostatistic estimation of time depths of the five main Chinese dialects]). *Gengo kenkyū* (言語研究) 38: 33–105.

Williams, Samuel Wells. 1856. *A Tonic Dictionary of the Chinese Language in the Canton Dialect*. Canton: Printed at the Office of the Chinese Repository.

Wu, Rui-wen (吳瑞文). 2014. Cóng bǐjiào Mǐnyǔ de guāndiǎn lùn Nánchāo Jiāngdōng fāngyán qí yùn de nǐcè (從比較閩語的觀點論南朝江東方言齊韻的擬測 [The reconstruction of the rimes of the Nánchāo Jiāngdōng dialect]). *Dōngfāng yǔyánxué* (東方語言學) 14: 60–80.

Wú, Wéi (吳維). 2016. Qīnzhōu Yuèyǔ yǔ Guǎngzhōu Yuèyǔ yǔyīn xìtǒng duìbǐ yánjiū (欽州粵語與廣州粵語語音系統對比研究 [Comparative Study of Qinzhou Cantonese and Guangzhou Cantonese Phonetic System]). *Guǎngxī jiàoyù xuéyuàn xuébào* (廣西教育學院學報) 1(2016): 45–47.

Xióng, Zhènghuī (熊正輝). 1987. Guǎngdōng fāngyán de qūfēn (廣東方言的區分 [The classification of Guǎngdōng dialects]). *Fāngyán* (方言) 5: 161–165.

Xióng, Zhènghuī (熊正輝), Zhāng Zhènxīng (張振興). 2008. Hànyǔ fāngyán de fēnqū (漢語方言的分區 [The classification of Chinese dialects]). *Fāngyán* (方言) 30(2): 97–108.

Xǔ, Jiéhóng (許潔紅). 2013. *Guǎngdōng Yángshān Yuèyǔ yǔyīn yánjiū* (廣東陽山粵語語音研究 [A phonetic study of Yangshan Cantonese]). Ph.D. diss., Jìnán University (Guǎngzhōu).

Xǔ, Wénxiàn (許文獻). 2001. *Zhànguó Chǔxì duō shēngfúzì yánjiū* (戰國楚系多聲符字研究 [Study of multi-reading characters of Warring States Chu]). Master's thesis, Department of Chinese Literature, Changhua Normal University.

Yáng, Huàndiǎn (楊煥典). 1990. Nánníng Yuèyǔ yǔyīn yǎnbiàn qūxiàng fēnxī (南寧粵語語音演變趨向分析 [Analysis of the Evolution of Cantonese Pronunciation in Nanning]). *Jìnán xuébào (zhéxué shèhuì kēxué bǎn)* (暨南學報哲學社會科學版) 1(1990): 79–84.

Yáng, Huàndiǎn (楊煥典), Liáng Zhènshì (梁振仕), Lǐ Pǔyīng (李譜英), Liú Cūnhàn (劉村漢). 1985. Guǎngxī de hànyǔ fāngyán (廣西的漢語方言 [The Chinese dialects of Guǎngxī]). *Fāngyán* (方言) 3(1985): 181–190.

Yáo, Yùmǐn (姚玉敏). 2009. Xiānggǎng Yuèyǔ shàngshēng biànhuà chūtàn: Yǔ yīn shíyàn yánjiū (香港粵語上聲變化初探: 語音實驗研究 [A Preliminary Study on the Change of Cantonese in Hong Kong: A Study of Speech Experiments]). *Yǔyán jì yǔyánxué* (語言暨語言學) 10(2): 269–291.

Yiu, T'ung. 1996. *The T'ai-Shan Dialect*. Ph.D. diss., Princeton University.

Yuán, Jiāhuá (袁家驊). 2001 [1960]. *Hànyǔ fāngyán gàiyào* (漢語方言概要 [Overview of Chinese dialects]). *Second Edition*. Běijīng: Yǔwén chūbǎnshè (語文出版社).

Yue-Hashimoto, Anne O. (余霭芹). 1973. *A Guide to the Tai-shan Dialect*. Washington, D.C.: Educational Resources Information Center.

Yue-Hashimoto, Anne O. (余霭芹). 1982. The differences between the literary and the vernacular readings in the Sui-hsi dialect. *Bulletin of the Institute of History and Philology*, Academia Sinica 53(2): 353–366.

Yue-Hashimoto, Anne O. (余靄芹). 1985. *The Suixi dialect of Leizhou: A study of its phonological, lexical and syntactic structure*. Ng Tor-tai Chinese Language Research Centre, Institute of Chinese Studies, Chinese University of Hong Kong.

Yue-Hashimoto, Anne O. (余靄芹). 1988. A preliminary investigation into the subclassification problem of the Yue dialects. *Computational Analyses of Asian and African Languages* 30: 7–41.

Yue-Hashimoto, Anne O. (余靄芹). 1991. Yuèyǔ fāngyán fēnqū wèntí chūtàn (粵語方言分區問題初探 [A Preliminary Study on the Subgrouping of Cantonese Dialects]). *Fāngyán* (方言) 3: 164–181.

Yue-Hashimoto, Anne O. (余靄芹). 1992. Guǎngdōng Kāipíng fāngyán de zhōngxìng wénjù (廣東開平方言的中性文句 [Neutral sentence of the Kaiping dialect of Canton]). *Zhōngguó yǔwén* (中國語文) 4: 279–286.

Yue-Hashimoto, Anne O. (余靄芹). 1993. *Comparative Chinese dialectal grammar: handbook for investigators*. Paris: Centre de Recherches linguistiques sur l'Asie orientale (École des hautes études en sciences sociales).

Yue-Hashimoto, Anne O. (余靄芹). 2002. Development of the stop endings in the Yue dialects. In: Dah-An Ho (ed.). *Dialect Variations in Chinese, Papers from the Third International Conference on Sinology, Linguistics Section*: 217–245. Taipei: Academia Sinica.

Yue-Hashimoto, Anne O. (余靄芹). 2006. *Táishān Dàncūn fāngyán yánjiū* (台山淡村方言研究 [The Dancun Dialect of Taishan]). Hong Kong: Language Information Sciences Research Centre, City University of Hong Kong.

Yue-Hashimoto, Anne O. (余靄芹). 2015. The Yue Language. In: William S.-Y. Wang, Chaofen Sun (eds.). *The Oxford Handbook of Chinese Linguistics*: 173–188. Oxford University Press.

Zēng, Jiànshēng (曾建生). 2014. Guǎngdōng Sìyì fāngyán gǔ qī, tòu, pāngmǔ cāyīnhuà (廣東四邑方言古溪, 透, 滂母擦音化 [Fricativization of kh, th, and ph initials in Sze Yap dialects].) *Yǔyán kēxué* (語言科學) 13(1): 96–104.

Zhān, Bóhuì (詹伯慧). 1981. *Xiàndài hànyǔ fāngyán* (现代汉语方言 [Modern Chinese Dialects]). Wǔhàn: Húběi rénmín chūbǎnshè (湖北人民出版社).

Zhān, Bóhuì (詹伯慧), Zhāng Rìshēng (張日昇). 1987. *Zhūjiāng sānjiǎozhōu fāngyán zìyīn duìzhào* (珠江三角洲方言字音對照 [Comparative pronunciation of Pearl River Delta dialects]). Guǎngzhōu: Guǎngdōng rénmín chūbǎnshè (廣東人民出版社).

Zhān, Bóhuì (詹伯慧), Cheung Yat-Shing (張日昇). 1988. *Zhūjiāng sānjiǎozhōu fāngyán cíhuì duìzhào* (珠江三角洲方言詞彙對照 [Comparison of vocabularies of Pearl River Delta dialects]). Guǎngzhōu: Guǎngdōng rénmín chūbǎn shè (廣東人民出版社).

Zhān, Bóhuì (詹伯慧), Zhāng Rìshēng (張日昇). 1990. *Zhūjiāng sānjiǎozhōu fāngyán zòngshù* (珠江三角洲方言綜述 [Summary of the Pearl River Delta dialects]). Guǎngzhōu: Guǎngdōng rénmín chūbǎnshè (廣東人民出版社).

Zhān, Bóhuì (詹伯慧), Zhāng Rìshēng (張日昇). 1994. *Yuè běi shíxiànshì Yuè fāngyán diàochá bàogào* (粵北十縣市粵方言調查報告 [Report of fieldwork in ten counties of northern Guangdong]). Jìnán dàxué chūbǎnshè (暨南大學出版社).

Zhān, Bóhuì (詹伯慧), Zhāng Rìshēng (張日昇). 1998. *Yuè xī shíxiànshì Yuè fāngyán diàochá bàogào* (粵西十縣市粵方言調查報告 [Report of fieldwork in ten counties of western Guangdong]). Jìnán dàxué chūbǎnshè (暨南大學出版社).

Zhāng, Jìngfēn (張靜芬). 2013. *Mǐnnán fāngyán de lìshǐ bǐjiào jí qí yǔyīn gòunǐ* (閩南方言的歷史比較及其語音構擬 [The historical comparison of the Minnan dialect and its phonetic reconstruction]). Master's thesis, Běijīng University.

Zhang, Meng-han, Pan Wuyun, Shi Yan, Li Jin. 2018. Phonemic evidence reveals interwoven evolution of Chinese dialects. Ms., available at: https://www.semanticscholar.org, accessed on May 2018.

Zhāng, Wèigāng (張為綱). 1943. Tàihé tòu dìng èr mǔ de liǎng zhǒng tèbié biànhuà zhī yánjiū (泰和透定二母的兩種特別變化之研究 [Special changes in tòu and dìng initials in the Tàihé dialect]). *Zhōngshān xuébào* (中山學報) 1(8): 72–82.

Zhào, Tóng (趙彤). 2006. *Zhànguó Chǔ fāngyán yīnxì* (戰國楚方言音系 [The sound system of the Warring States Chu dialect]). Běijīng: Zhōngguó xìjù chūbǎnshè (中國戲劇出版社).

Zhōu, Zhènhè (周振鶴), Yóu Rǔjié (遊汝傑). 1985. Húnán shěng fāngyán qūhuà jí qí lìshǐ bèijǐng (湖南省方言區劃及其歷史背景 [The classification of Húnán dialects and their historical background]). *Fāngyán* (方言) 4(1985): 257–272.

Appendix I. Developments of initial consonants in Yue dialects.

dialects	1	2	3a	3b	4	5a	5b	6a	6b	7a	7b	8	9	10	11a	11b	11c
Guǎngzhōu (廣州)																	+
Huāxiàn (花縣)																	+
Cónghuà (從化)													+				+
Xìnyí (信宜)													+			+	
Gāozhōu (高州)													+			+	
Nánnìng báihuà (南寧白話)														+			+
Fóshān (佛山)																	+
Nánlǐng jiǔjiāng (南嶺九江)			+				+					+			+		
Nánhǎi shātān (南海沙灘)	+					+											+
Shùndé dàliáng (順德大良)	+				+	+						+					+
Sānshuǐ (三水)												+					+
Gāoyào (高要)	+		+														+
Wúzhōu (吳州)		+				+				+				+	+		
Huàxiàn (化縣)		+				+				+			+	+		+	
Cāngwú (蒼梧)	+				+	+							+	+		+	
Téngxiàn (藤縣)	+					+				+		+		+		+	
Róngxiàn (容縣)	+			+		+				+				+		+	
Yùlín (玉林)	+			+		+				+				+		+	
Shínán (石南)	+			+										+		+	
Bīnyáng (賓陽)	+			+										+		+	
Nānnìng pínghuà (南寧平話)	+													+		+	

	Dōngwǎn (東莞)		+			+											+
	Bǎo'ān (寶安)																+
	Qīnzhōu (欽州)													+		+	
	Zhōngshān (中山)												+		+		
	Zhūhǎi (珠海)												+		+		+
	Jiāngmén (江門)									+							+
五邑	Xīnhuìchéng (新會城)					+				+				+		+	
	Xīnhuìhécùn (新會河村)									+				+		+	
	Táishān (台山)						+		+					+	+	+	
	Kāipíng (開平)						+		+			+		+	+	+	
	Ēnpíng (恩平)									+				+		+	
	Hèshān (鶴山)						+		+			+			+		+
	Dǒuménzhèn (斗門鎮)							+		+						+	
	Yángjiāng (陽江)													+	+		+
	Yángchūn (陽春)													+	+		+
	Sīhè (思賀)														+		+
	Guīpíng (桂平)					+								+	+		+
	Píngnán (平南)					+									+		+

Data taken, with minor revision, from Yue-Hashimoto 1991.

List of features

Number	description
1	medieval Chinese voiced consonants > tenuis
2	medieval Chinese voiced consonants > voiceless aspirated
3a	medieval Chinese *k- becomes palatalised, only in Grade III rimes of traditional *liú shè* 流攝, *shēn shè* 深攝, *zhēn shè* 臻攝

3b	medieval Chinese *k- becomes palatalised, only in Grade III rimes of traditional *liú shè* 流攝, *shēn shè* 深攝
4	medieval Chinese *jīng zǔ* 精組 (*ts-, *ts^{h}-, *dz-, *s-, *z-) > ts-, ts-, s-; *zhī zǔ* 知組 (*ʈ-, *ʈ^{h}-, *ɖ-, *ɳ-) and *zhào zǔ* 照組 (*tʂ-, *tʂ^{h}-, *dʐ-, *ʂ-, *ʐ-) > tɕ-, tɕ^{h}-, ɕ-
5a	medieval Chinese *jīng zǔ* 精組 (*ts-, *ts^{h}-, *dz-, *s-, *z-) > t-, t^{h}-
5b	medieval Chinese *jīng zǔ* 精組 (*ts-, *ts^{h}-, *dz-, *s-, *z-) > t^{h}-
6a	medieval Chinese *t- > 0
6b	medieval Chinese *t^{h}- > h
7a	medieval Chinese *p- > b-, *t- > d-
7b	medieval Chinese *p- > v-
8	medieval Chinese *ɦ- > h-
9	medieval Chinese *k$^{(w)}$- and *k$^{h(w)}$- have merged
10	medieval Chinese *s- (in some dialects also *dz-, *z-, *ɕ-) > ɬ- or θ
11a	medieval Chinese *ŋ- > ŋ-
11b	medieval Chinese *ŋ- > ɳ-
11c	medieval Chinese *ŋ- > j-

It is interesting to note that there is no other branch, except for Ng Yap, which would exhibit features that are not shared by any other Yue dialect. Features 6a and 7a are unique distinctions of the Ng Yap branch. Features 4, 5ab, in the opinion of the present writer, are highly speculative; there is no real evidence that these phonemic mergers have really occurred.

Джорджио Орланди. Классификация диалектов нг-яп: к вопросу о принципах выделения подгрупп внутри синитической семьи

Диалекты группы нг-яп (ранее — сэй-яп) обычно считаются подветвью диалектной группы юэ. В настоящей статье предпринята попытка показать, что, несмотря на это широко распространенное мнение, диалекты нг-яп обнаруживают целый ряд дистинктивных особенностей, которые с точки зрения формальной языковой классификации вынуждают нас обособить их от группы юэ. В статье также обсуждаются общие критерии, используемые для диалектной классификации в современной синологии. Работа ориентирована в первую очередь на то, чтобы стимулировать дальнейшую дискуссию по данной теме, долгое время игнорировавшейся в китайской диалектологии.

Ключевые слова: диалекты нг-яп, диалекты юэ, языковая классификация, синитические языки, китайские диалекты.

George STAROSTIN
Higher School of Economics / Russian State University for the Humanities / Santa Fe Institute; gstarst1@gmail.com

Chinese basic lexicon from a diachronic perspective: implications for lexicostatistics and glottochronology[1]

In this paper, I attempt to compare the relative rates of replacement of basic vocabulary items (from the 100-item Swadesh list) over four specific checkpoints in the history of the Chinese language: Early Old Chinese (as represented by documents such as *The Book of Songs*), Classic Old Chinese, Late Middle Chinese (represented by the language of *The Record of Linji*), and Modern Chinese. After a concise explication of the applied methodology and a detailed presentation of the data, it is shown that the average rates of replacement between each of these checkpoints do not significantly deviate from each other and are generally compatible with the classic «Swadesh constant» of 0.14 loss per millennium; furthermore, these results correlate with other similar observed situations, e.g. for the Greek language, though not with others (Icelandic). It is hoped that future similar studies on the lexical evolution of languages with attested written histories will allow to place these observations into a more significant context.

Keywords: Chinese language history, Old Chinese language, Middle Chinese language, lexicostatistics, glottochronology, basic vocabulary.

Introduction

Over the last couple of decades, lexicostatistical methodology has played an important role in historical studies on the evolution of various «dialectal» forms of Chinese (or «Sinitic languages», from a more strictly linguistic point of view). Since there is no universally accepted model of the lexicostatistical procedure as far as the selection of source data, manual and/or automated annotation of lexical cognates, and the specific phylogenetic algorithm applied to the data are concerned, these studies significantly vary in terms of selected scope, stated goals, and attained results; but there seems to be a general understanding that conducting lexicostatistical studies is an important stage in unraveling the internal history of Chinese and identifying certain key points resulting in divergent linguistic lineages, as well as separating evidence for genetic splits from evidence for later linguistic contacts that tend to obscure the different lineages in question.

That said, most of the studies on Chinese (Sinitic) lexicostatistics have largely focused on quantifying and interpreting the degree of lexical divergence between modern colloquial forms of Chinese[2], usually downplaying the important fact that Chinese is one of the very few

[1] I thank Prof. Laurent Sagart for his valuable comments on parts of this paper, and Dr. Johann-Mattis List for the opportunity to present its major points before a large audience of specialists at the *Old Chinese And Friends* conference (Max Planck Institute for the Science of Human history, Jena, April 26–27, 2018). Any errors in data or its analysis are exclusively my own.

[2] It is not within the scope of the current paper to provide a detailed listing of all the works that have applied quantitative methods to the problem of Chinese dialect classification. For those unfamiliar with the topic a good starting point could be the complex study of Mahé Ben Hamed and Wang Feng (2006), who apply a variety of dis-

languages in the world whose historical evolution can actually be studied by means of preserved written data, rather than reconstructed through the comparative method — and, consequently, one of the most important test cases in the world (along with several Indo-European and Semitic languages) when it comes to measuring rates of lexical evolution[3].

The reasons for such negligence are understandable. Studying lexical replacement in languages represented only by a closed and limited corpus of written data necessarily runs into a number of uncertainties — insufficient attestation of required items in available texts, their occasional semantic ambiguity, and lack of direct knowledge on the dialectal characteristics of said texts, among other things. To make matters worse, historically attested forms of Chinese are commonly understood to mix together different strata — to the point that, for instance, our current understanding of Middle Chinese phonology (as extracted from rhyme books and rhyme tables) vastly exceeds our understanding of Middle Chinese grammar *and* lexicon, since most texts in the classic era of Táng and early Sòng dynasties were written in one or another variant of the archaic Literary Chinese. Circumstances such as these may seem to make the painstaking task of studying lexical replacement within Chinese in detail a waste of time, but in reality it is not that difficult to employ a somewhat formalistic approach to the matter and at least try to see what it gets us. However, in order for such a study to be of any use, it is imperative to state the rules very clearly and consistently apply them to all selected time periods and data collections.

The present paper is a tentative attempt to manually measure the rates of lexical evolution over a period of approximately 2,500–2,800 years in the history of Chinese. This is achieved by selecting several chronological checkpoints, constructing standardized Swadesh wordlists for each of them and individually investigating each certified or potential case of lexical replacement from one checkpoint to another. Two reasons why such a study, though still clearly far from perfect, could not have appeared earlier, are as follows: (a) a breakthrough in corpus studies on Old Chinese — largely due to the outstanding dedicated work of Donald Sturgeon and his colleagues, we now have the advantage of the online *Chinese Text Project*, allowing for complex lexical investigations on a large scale to be conducted almost momentarily; (b) significant methodological clarification of the lexicostatistical technique, described in several papers from the Moscow school of comparative linguistics (see the "Methodology" section below). Naturally, there is still much room for improvement (especially in the area of Middle Chinese, which remains considerably underdeveloped), but there is reason to believe that even at this stage, the results will be useful enough both for Sinologists and general specialists in diachronic linguistics.

Before presenting the data in its entirety, it is necessary to do the following things: (a) justify and describe the four selected chronological checkpoints — Early Old Chinese, Classical

tance- and character-based methods in order to determine whether the configuration of known forms of Chinese better agrees with a tree-like or a network-like structure; the same data was later made use of by Johann-Mattis List (2015) in his own investigation of the historical relations between Chinese dialects. Further references to earlier studies may be easily found in either of those papers.

[3] To the best of my knowledge, only two brief attempts at measuring the lexical distance between Old Chinese and Modern Chinese have had their results mentioned in literature: (a) Swadesh 1952: 456 and subsequent papers by both Swadesh and Robert Lees make frequent reference to the results of C. Y. Fang, who allegedly established that 79% out of the 200-item wordlist of «Classic Chinese 950 A.D.» have been retained in «modern colloquial Northern Chinese»; the wordlist itself has never been published, making it impossible to verify the claim, and it is in fact quite unclear what is meant by «Classic Chinese 950 A.D.»; (b) Starostin 2000: 256 actually gives a specific list of replacements from «Archaic Chinese (seventh century BC)» to modern Mandarin that can be checked, and the verification shows a significant number of omissions (see below for specific examples).

Old Chinese, (Late) Middle Chinese, and Modern Chinese, including some discussion on dating, data sources, and various technical problems; (b) give a brief description of the methodology employed in selecting items for the respective positions in the wordlist, as well as the procedure of cognate scoring from one period to another. This will be followed by reasonably detailed discussion of the data itself, after which we present a brief analysis and state our conclusions on the tendencies of lexical evolution in the history of Chinese, including a typological-comparative angle.

Data sources

1. Early Old Chinese (EOC)

Definition: we approximately define Early Old Chinese as the language that is represented in writing by such literary monuments as the *Shījīng* ('Book of Odes') and the oldest parts of the *Shūjīng*, or *Shàngshū* ('Book of Documents'), as well as epigraphic data from artefacts (mainly bronze vessels) dating back to the Early Zhōu dynasty (*jīnwén*); the most comprehensive and systematic Western dictionary of this language is Schuessler 1987. In general, the language of all these texts is known to share certain grammatical and lexical properties that strongly distinguish it from later forms of Chinese, though it cannot be said for certain to represent a direct ancestral stage for any of them.

Reasons for selection: EOC is the very first chronological checkpoint for which it is possible to construct anything close to a standardized Swadesh wordlist. Although some observations may be made on certain elements of the basic lexicon in the oracle bone inscriptions of the Shāng dynasty, the restricted and highly formulaic nature of these inscriptions leads to way too many gaps in the wordlist for it to be of any use for the present study. Therefore, a general statistically relevant investigation of Chinese basic lexicon may only begin from Early Zhōu times.

Sources: Much, if not most, of the epigraphic material is ineligible for the task of building a Swadesh wordlist due (once again) to the highly formulaic subject matter and ritualistic nature of the texts, leaving the verses of the *Shījīng* as the single most natural source for an EOC list of basic lexicon. Out of the 100 required items, only eight ('ashes', 'bark', 'bone', 'egg', 'knee', 'lie', 'liver', 'louse') have no reliable or probable equivalents attested in the *Shījīng* (or in the eldest parts of the *Shūjīng*).

Problems: There is little doubt that the texts of the *Shījīng* are relatively heterogeneous in terms of both time and space (see Dobson 1968: 224–242 for an attempt at a chronological linguistic stratification of the various sections of the *Shījīng* based on grammatical evidence), but there is so far very little evidence that the dialects of the *Shījīng* were significantly different from each other as far as their basic lexicon was concerned: very few synonyms for basic notions were elicited from the data, and those that were elicited are not easily described in terms of dialectal variety (see, e.g., 'give' below, with two different synonyms attested in the exact same poem). A much more significant problem is the scarceness of attestation for multiple terms: in many cases unambiguous contexts with the required word are found but once or twice, and their reliability often depends on external data (e.g. if the same word is also the basic equivalent for the term in Classical Chinese, this improves the chances of the corresponding item in EOC). All such terms are specially commented upon in the notes section, and particularly dubious inclusions are marked with a question sign.

2. Classical Old Chinese (COC)

Definition: We define COC as the language of literary texts, most likely reasonably close to the spoken language of the time, written from approximately the end of the 5th century to the end of the 3rd century BC. There is no single defining dictionary for this stage of the language, since lexicographical sources usually conflate it either with EOC or with Hàn-era OC (or both); however, the text corpus is reasonably well defined, and focused searches may be performed these days with the aid of such resources as the *Chinese Text Project* (Sturgeon 2019).

Reasons for selection: COC is the first known historical stage of Chinese that is represented by a substantial amount of thematically diverse non-poetic texts which, according to a general scholarly consensus, are written in a language that reasonably closely reflects colloquial patterns of the time (with certain expected stylistic emendations, though their influence on core basic lexicon is probably negligible). A significant advantage of this period is that the language of the texts in question is not as highly influenced by the language of the previous period (EOC) as the written language of Hàn-era and later periods is dependent on COC.

Sources: COC is generally understood to have possessed a significant amount of dialectal diversity; even if evidence for this rarely comes from core basic vocabulary, for the sake of increased accuracy we prefer to draw upon sources typically recognized to stem from the same dialectal area. The principal texts corroborating our selections are the *Lùn yǔ* and (especially) *Mèng-zǐ*, both recognized as representative of the Lǔ dialect (Pulleyblank 1995: 3), although there may be a chronological gap of about 100–150 years in their original composition (not essential for our purposes).

If the necessary words are encountered very rarely or not encountered at all in these texts, we find it acceptable to draw upon data from other sources, such as the *Zuǒ zhuàn* (representing a separate dialect of its own, together with the *Guó yǔ*) and *Zhuāng-zǐ* (probably representing a more Southern, Chǔ-area, dialect, though this is debatable). For the record, the following words are not attested at all in the *Lùn yǔ* and *Mèng-zǐ* and have to be substituted from other sources: 'ashes', 'bite', 'nail', 'dry', 'green', 'knee', 'liver', 'louse', 'red', 'root', 'round', 'sand', 'smoke', 'swim', 'tail', 'tongue'. Since every single one of these 16 items is either the same as in EOC or the same as in MC or both, and since we have been unable to reliably elicit even a single undeniable difference in the Swadesh wordlist between any of the listed texts, such substitution should be permissible.

Problems: COC is (arguably) one of the least problematic periods in the history of Chinese when it comes to eliciting basic lexicon; see above on the relative insignificance of dialectal divisions for this purpose. Several dubious cases of elicitation, usually having to do with scarceness of attestation and ambiguity of translation, are commented upon specifically in the data section of the paper.

3. Middle Chinese (MC)

Definition: For the purposes of the current study, Middle Chinese is narrowly defined as the colloquial (or reasonably close to colloquial) stage of Chinese, chronologically coinciding with or closely following the beginning of the division of Chinese into the principal dialectal groups of today, i.e. what is commonly called *Late* rather than *Early* Middle Chinese. This is due to the fact that texts from the Early Middle Chinese era (first half of the Táng dynasty, 7th–8th centuries AD) are nearly always written in an archaic form of the language (*wén yán* or *gǔ wén*), whereas for the Late Middle Chinese period (late Táng and early Sòng dynasties) there is a limited, but useful corpus of textual evidence that is somewhat sufficient for purposes of lexicostatistical analysis.

Reasons for selection: The entire period between COC and the 20th century is an extremely difficult area for lexicostatistical evaluation, since almost every text written in traditional imperial China, from Hàn all the way to Qíng dynasties, is influenced, to various degrees, by the grammar and lexicon of COC, and hardly ever reflects the spoken language of the corresponding period. It is precisely for this reason that we have refrained, for instance, from attempting to construct a separate 100-item wordlist for the language of the early or late Hàn dynasty, despite the abundance of textual evidence from that period — perusal of such vast sources as Sīmǎ Qiān's *Shǐ jì*, for instance, shows that in many cases Swadesh items are represented by at least two competing equivalents (e.g. 犬 *quǎn* and 狗 *gǒu* for 'dog', 盈 *yíng* and 滿 *mǎn* for 'full', etc.), and it is often impossible to determine whether such situations are due to true «transit synonymy» (when a lexical innovation has not yet fully managed to displace the original neutral term) or to the intentional (or unintentional) mixing of standard colloquial and outdated archaic equivalents.

Nevertheless, it is very important to have at least one analyzable «checkpoint» on the almost 2,500 year long way from COC to Modern Chinese, and from a general chronological and qualitative point of view, Late Middle Chinese is the optimal, if far from perfect, candidate for this purpose, since this is the period of proliferation for the genre of the 語錄 *yǔlù* («records of sayings»), a new genre of Buddhist literature whose innovative and frequently iconoclastic nature placed a large emphasis on transmitting sermons, parables, and real life anecdotes by means of colloquial idioms. In general, the *yǔlù* may be considered as the first fully colloquial genre of literature in the history of past-COC Chinese, and although it is more thematically limited than the fictional genres of late Sòng, Yuán, and Míng dynasties, its advantages are that it is represented by chronologically older texts and that at least some of these texts are arguably more free from literary embellishments than the literary genres of *huàběn* and *xiǎoshuō* (classic short stories and novels from Sòng to Míng-Qíng times).

Sources: A thorough lexical analysis of all or most of the existing texts in the *yǔlù* genre has not been conducted yet; an important problem is that some of the texts may reflect serious dialectal differentiations. For this reason, analysis has so far been restricted to just one reasonably large and generally uniform specimen of the genre, namely, the *Línjì yǔlù* («The record of Linji»), a text traditionally attributed to the disciples of the school of Master Linji Yixuan (d. 866 AD) but not finalized until the late 11th–early 12th centuries. The language of the *Línjì yǔlù* and the *yǔlù* genre has been the subject of several meticulous studies, e.g. Sawer 1969, Gurevich 2001, but all of them focus almost exclusively on grammar rather than lexicon; nevertheless, analysis of the basic words used in the text is in perfect agreement with the grammatical data in that the *Lǐnjì yǔlù* does indeed attempt to represent the colloquial standards of its time, albeit with some inescapable influence of the more classical forms as well.

Problems: Restriction to a single source necessarily implies that our MC list has the heaviest gaps of all (at least 18 out of 100 items are not featured at all in the text, and 8 more are somewhat problematic due to scarceness of attestation and semantic ambiguity); the problem is somewhat alleviated by the fact that the majority of these gaps are items that are represented by the same equivalent in COC and Modern Chinese, so it may be reasonably assumed that they were not replaced by anything else in MC as well.

4. Modern Chinese (PTH)

Definition: Since this study is only concerned with the issue of relatively straightforward diachronic evolution from a single point in the past to a single point in the present, we intentionally limit our definition of «Modern Chinese» to the present day version of *pǔtōnghuà*, the

common national language generally based on the Běijīng Mandarin dialect; linguistic differences between the actual spoken varieties of Běijīng Mandarin and *pǔtōnghuà* are well known, but do not generally extend to core basic vocabulary, making this factor negligible.

Reasons for selection: Theoretically, any other Chinese «dialect» (with the exception of Mǐn, since that cluster is typically assumed to have split off from the rest before the beginning of the MC period) might have been substituted here, but the task of constructing a 100-item wordlist for *pǔtōnghuà* is naturally easier than for any of the rest. A separate study is necessary to assess the rate of evolution from MC to PTH relative to other varieties of spoken Chinese that are in use today.

Sources: A variety of sources has been used (textbooks, dictionaries, text corpora, live informants etc.).

Problems: This is the least problematic area of all; issues are typically limited to purely technical problems, such as choosing a monosyllabic or bisyllabic variant for the most common equivalent of a given meaning (where the adopted solution usually bears no impact on calculations anyway).

Methodology of wordlist construction and lexical comparison

In constructing the optimal wordlists for each of the four stages, I attempt to follow as closely as possible the guidelines laid down in Starostin 2010 and Kassian et al. 2010, which can largely be boiled down to the following principles: (a) elicit words whose meaning and stylistic register are as close as possible to the pre-defined meanings listed in the latter paper; (b) try to avoid the inclusion of multiple synonyms, whose presence undermines the main idea of lexicostatistics.

Obviously, when dealing with written stages of the language represented by closed (and usually not very large) corpora, formal and precise adherence to these principles is not always possible. Due to the nature of the data itself, all of the wordlists presented below, with the exception of the wordlist for Modern Chinese, will inevitably contain errors, some of which might not even be rectified in the future unless massive new amounts of data (e.g. from archaeological sources) become available. However, the important thing here is to make certain that these errors do not skew the quantitative conclusions in any one particular direction, i.e. that they do not increase specifically the number of lexical replacements or the number of lexical retentions from any chosen point in the history of Chinese to the next one. This implies the necessity of a transparent, objective, well-argued methodology of dealing with ambiguous situations, one that should preferably minimize the possible interference of the personal preferences of the compiler. Below I list some of the general points; specific applications may be found in the comments on particularly troublesome lexical items in their respective sections.

1) *Be wary of etymological arguments*. Frequently, when facing the choice between picking one out of two or more synonyms, or including all of them into the list, one may be led astray by the fact that an older equivalent, clearly going back to the original main equivalent for a given Swadesh term, is still preserved at a later stage in the development of the language — ignoring the clear fact that its semantics has shifted, as the word is now used in a slightly different meaning or has been relocated to a different (marked) stylistic register (vulgar or elevated).

This is, for instance, the reason of several important mistakes in Starostin 2000: 256, a general study in the methodology of lexicostatistics where Old Chinese is compared with Modern Chinese and 23 lexical replacements are identified. The study fails to list several transparent

replacements, such as 目 *mù* → 眼睛 *yǎn-jing* 'eye' and 首 *shǒu* → 頭 *tóu* 'head', presumably because the former equivalents are still encountered today in various bound idiomatic formations and archaic contexts. This leads to underestimations of the process of lexical replacement, and the problem gets even worse for eras that are only represented by written documents, since written language by its very nature fails to keep up the pace with developments in the colloquial idiom, and special care must be given to the study of preserved texts in order to make a qualified decision on whether a certain lexical replacement has already been completed at a given period or not. In any case, 'etymological argument' alone, not supported by actual data from texts, does not carry significant value.

2) *Watch out for bound forms and idiomatization*. The «basic» equivalent of any given meaning is typically understood as the most neutral and generally context-independent form: the more words there are that an observed candidate can enter in syntactic relations with, the better are its chances for historical stability. Thus, COC has multiple equivalents for the meanings 'die' and 'kill', but a great majority of them has limited syntactic applicability: e.g. 弑 *shì* 'to kill' is only used in reference to killing a superior (prince, father, etc.), whereas 薨 *hōng* 'to die' is only said of high officials. Not surprisingly, these are precisely the words that do not survive into the MC era, whereas the neutral 殺 *shā* 'kill' and 死 *sǐ* 'die' persist all the way into most modern forms of Chinese.

3) *Textual evidence is generally superior to dictionary information*. With a closed and relatively limited textual corpus that is not particularly well reflected in specialized dictionaries, OC is clearly one of those ancient languages where direct elicitation of lexical data from the corpus is much preferable to relying on dictionaries. In a few cases, observations of actual word usage in the attested texts may lead to startlingly unpredictable conclusions (see notes on possible replacements from EOC to COC below); more importantly, finding relevant syntactic and semantic contexts adds a much wanted level of confidence to our wordlists, and also helps differentiate between statistically frequent and rarely used synonyms. This is particularly helpful for transitional stages of the language, in which an older equivalent may already be retained only as a rare archaism (including quotations from and paraphrasing of older texts), while the newer replacement may be more frequent — however, such situations will rarely, if ever, be discussed or even hinted at in dictionaries.

Regarding the procedure of cognate scoring, in this particular setting it is essentially reduced to the procedure of *postulating lexical replacements from one time period to another*. In addition to the obvious (lexical replacements are assumed whenever word X, used in a given Swadesh meaning over the time period t_n, is no longer used in that meaning over the time period t_{n+1}), we try to observe the following rules:

1) *Statistics and stylistics matter*. This is essentially a recapitulation of points 1 and 2 from the previous section: even if the same word is encountered seemingly in the same meaning over several distinct time periods, this does not always imply that it has not actually been replaced. Written Chinese has always operated according to the «forget nothing» principle: no matter how archaic a certain word is, there is always some probability of encountering it in texts that are separated by any number of years from its time of proliferation. What matters is primarily the statistics of usage (if there are two or more synonyms, which one is the most frequent?) and the stylistic context of usage (if there are two or more synonyms, which ones are used in quotations, poetic formulas, imitations of archaic rhetorics — and which ones are used in colloquial direct speech or neutral descriptions of situations?). If it can be shown that synonyms A and B express the same meaning in t_{n+1} as exclusively A in t_n, but that A is rare compared with B *and* primarily used in stylistically marked contexts, we postulate a clear-cut lexical replacement.

2) *Morphological change does not matter*. The issue of «partial cognacy», where two equivalents of the same Swadesh meaning in two different languages (or different stages of the same language) consist of two or more morphemes, of which only one (usually the root) is etymologically shared between them, while the others are different, seems to be particularly acute for languages that frequently resort to compounding techniques, including Chinese. This issue has been discussed several times in literature (e.g. List 2016; Starostin 2013a), but still remains without a perfect solution. Should a difference such as COC 知 *zhī* 'to know' vs. Modern Chinese 知道 *zhī-dào* id. be reflected by assigning both items the same index of cognacy (no lexical replacement), different indexes (replacement), or marked in some other manner (e.g.awarded «half a point» instead of a regular full +1 index, etc.)?

In my opinion, a definitive solution to this issue is impossible until a solid experimental base for this type of situations has been built up — which would allow us to cross-linguistically compare replacement rates for different methods of scoring and choose the solution that would make more general sense from a historical point of view. In the meantime, for Chinese I prefer to stick to the «no lexical replacement for partial cognacy situations» scenario, for the following reason: in most cases, morphemic compounding in the history of Chinese is explainable by reasons that have nothing to do with semantic shifts and more to do with the phonetic evolution of the language (avoidance of ever-increasing levels of homonymy), which is clearly not what we really want to measure when choosing lexical change as a base parameter for glottochronology. Therefore, in this study classical 知 *zhī* will be scored exactly the same as modern 知道 *zhī-dào*.

However, one important thing about both classical and modern Chinese compounds («binomes») that should be stated is that in many (not all) cases a binome may easily be analyzed as containing a *primary* and a *secondary* morpheme. The *primary* morpheme is the historical root morpheme; its defining diachronic characteristic is that it tends to be more stable over both time and space, and its defining synchronic characteristic is that, unlike the secondary morpheme, it can still be frequently encountered, usually in bound form, without the secondary morpheme in its original meaning. The *secondary* morpheme largely acts as an additional determiner: as a rule, it is less stable across time periods and dialects, it may be omitted in certain contexts, and whenever encountered on its own, it is rarely or never used in the same meaning as the primary morpheme.

A good example is Modern Chinese 月亮 *yuè-liàng* 'moon', where 月 *yuè* is the primary morpheme because it may be encountered on its own in the same meaning (usually in other bound forms, e.g. 月夜 *yuè-yè* 'moonlit night, etc.), whereas 亮 *liàng* 'light, shine' is never encountered with the meaning 'moon' if not in conjunction with 月 *yuè*; not surprisingly, 月 *yuè* is also the historically stable morpheme 'moon', common for most varieties of Chinese, whereas 亮 *liàng* is a more recent addition and alternates with other additions in different dialects (e.g. 月光 *yuè-guāng*, 月子 *yuè-zi* etc.).

Somewhat more complicated are cases of concatenated binomes in which, upon first sight, both morphemes express the same meaning and are hard to classify as respectively primary and secondary — such as 道路 *dào-lù* 'road' (literally 'road$_1$' + 'road$_2$') or 牙齒 *yá-chǐ* 'teeth' ('tooth$_1$' + 'tooth$_2$'). It would seem that technically, the best solution here would be to judge the two morphemes as synonymous and include both into the calculations. However, even in this situation analysis of the behavior of the respective meaning in different contexts actually shows that one morpheme typically prevails over the other. Thus, in the meaning 'road' Modern Chinese frequently employs simple 路 *lù* (大路 *dà lù* 'big road', etc.), but practically never 道 *dào* (which is far more common in the abstract meaning 'way, manner'); the meaning 'tooth / teeth' is frequently expressed by 牙 *yá* (as in 刷牙 *shuā yá* 'brush one's teeth', etc.), but almost

never by 齒 *chǐ*. I interpret this as clear evidence that in forms such as 道路 *dào-lù* and 牙齒 *yá-chǐ*, one morpheme still behaves as primary and the other one as secondary, even if from a historical point of view (as can be seen from comparison with OC evidence, see the data below) it is the secondary morpheme that reflects the original Swadesh equivalent — see, however, the «be wary of etymological arguments» point above, which clearly pressures us into regarding such situations as lexical replacements.

One might argue that such a solution directly contradicts the «morphological change does not matter», but this is only if we understand the dynamic genesis of such compounds as 牙齒 *yá-chǐ* as the extension of the primary morpheme 齒 *chǐ* with the «prefixed» quasi-synonymous morpheme 牙 *yá*, when in reality the process must have been far more complex: equivalents of the monosyllabic 牙 *yá* are found in the basic meaning 'tooth' in many Chinese dialects, as well as alternate binomes such as 牙巴 *yá-ba*, etc., indicating that the structure of 牙齒 *yá-chǐ* is, in fact, quite analogous to that of 月亮 *yuè-liàng*. Ignoring this would mean ignoring an important element of lexical restructuring in the history of Chinese, and while other formal solutions are possible, in this study we will try to consistently apply this principle to the procedure of cognate scoring.

Notes on transcription

Since this study is only concerned with different stages of Chinese and not with the Sino-Tibetan (or areal) origins of the Chinese entries, issues of phonetic and phonological reconstruction of OC and MC are largely irrelevant; cognate identification is not required between OC, MC, and PTH, and phonological or phonetic transcriptions of Chinese characters only matter inasmuch as the paper might also interest general historical linguists with no knowledge of Chinese hieroglyphics, or, occasionally, to specify which particular pronunciation out of several possible ones is meant for a specific character (e.g. 長 **draŋ* > *cháng* 'long', not 長 **traŋʔ* > *zhǎng* 'grown-up', etc.).

Throughout the study, I consistently use the OC reconstruction of Sergei Starostin (1989), some of the aspects of which remain controversial (e.g. the reconstruction of lateral affricates and voiced aspirates, or the interpretation of Type A / Type B syllable distinction as reflecting an opposition in vowel length) but which I also find reasonably conservative in comparison with the far-reaching changes in Baxter, Sagart 2014. OC Reconstructions are taken either directly from Starostin 1989 or from Sergei Starostin's unfinished etymological database on Old Chinese («Chinese Characters Database» at the Tower of Babel website, http://starling.rinet.ru). MC readings are used very sparsely throughout the rest of the paper; where necessary, they are also taken from Starostin's database. Modern Chinese forms are transcribed in standard *pinyin*. OC and modern readings are typically given back-to-back next to the respective characters, with OC reconstructions accompanied by asterisks.

The data

All four wordlists have been published online as part of the Sinitic 100-item wordlist database, included in the Global Lexicostatistical Database framework (http://starling.rinet.ru/new100); in addition to the words themselves, the database includes plenty of annotations and comments, such as precise references to sources, quotations of contexts from which the items have been elicited, and (sometimes highly detailed) explanations on why certain synonyms were pre-

ferred over others. This section of the paper represents a seriously condensed, but also partially reworked variant of that part of the database, with all the words rearranged in order of their relative historical stability.

First I discuss the subset of «super-stable items» that have been retained from EOC all the way to PTH (this is the largest sub-set, but also understandably requiring the least amount of commentary); then group B consists of «medium-stable items», for which it makes sense to postulate one replacement over the analyzed 2,500-year long period; finally, the shortest and the most difficult group C consists of «highly unstable» items that may have undergone no fewer than two replacements over the same period. Group D lists two interesting deviations where intermediate periods may show «dead-end» dialectal semantic developments, and, finally, Group E lists one item that has been excluded from analysis due to insufficient data.

A. Super-stable items (61 words).

A.1. Items attested with the same root morpheme throughout all four stages of Chinese.

A.1.1. 'big': 大 (**dhaːts* > *dà*).

A.1.2. 'black': 黑 (**s=məːk* > *hēi*). ◊ Transparently derived from 墨 **məːk* > *mò* 'ink', but still clearly the primary neutral equivalent for 'black' already in EOC. The idea that 黑 *hēi* had replaced the earlier 玄 **gʷiːn* > *xuán* in this meaning during the Zhōu period (Schuessler 2007: 277) seems to rely more on the derived origin of *hēi* than concrete textual evidence: there are, in fact, no contexts at all in EOC or COC literary monuments where *xuán* should be unambiguously translated as 'black' rather than a more general 'dark'[4]. For a good context supporting a basic function for 黑 *hēi* (as well as 赤 *chì* 'red', see below), cf. 莫赤匪狐莫黑匪烏 *mò chì fěi hú, mò hēi fěi wū* «there is nothing redder than a fox, nothing blacker than a raven» (*Shījīng* 41, 3); no such diagnostic contexts are available for 玄 *xuán* or any of the even more rare quasi-synonyms for 'black, dark', such as 緇 *zī* (only found twice in the *Shījīng* applied to some names for garments).

A.1.3. 'blood': 血 (**swiːt* > *xuè*).

A.1.4. 'cloud': 雲 (**whən* > *yún*).

A.1.5. 'come': 來 (**rəː* > *lái*).

A.1.6. 'die': 死 (**siyʔ* > *sǐ*).

A.1.7. 'dry': 乾 (**ghar* > *gān*).

A.1.8. 'ear': 耳 (**nhəʔ* > *ěr*). ◊ In the modern language, used primarily as part of the binome 耳朵 *ěr-duǒ*, lit. 'ear-cluster'.

A.1.9. 'fire': 火 (**səməːyʔ* > *huǒ*).

A.1.10. 'fish': 魚 (**ŋha* > *yú*).

A.1.11. 'hair /of head/': 髮 (**pat* > *fá*). ◊ All four stages of Chinese show a very clear and persistent lexical differentiation between **pat* 'hair of the head' (in the modern language, typically used as part of the binome 頭髮 *tóu-fá* 'head-hair') and 毛 **mhaːw* 'hair on the body' (also 'wool', 'fur', etc.).

A.1.12. 'hand': 手 (**tlhuʔ* > *shǒu*).

[4] A different opinion is voiced in Wu 2011: 87, where it is stated that in the corpus of bronze inscriptions, 玄 *xuán* is more frequent than 黑 *hēi* and is a better candidate for «basic 'black'» than the latter. However, Wu does not list any diagnostic contexts; frequency alone is not a clinching argument here, if, for instance, 玄 *xuán* (like 緇 *zī* in later received texts) was a typical term for denoting specific shades of ceremonial clothing, frequently depicted in bronze inscriptions. Note that most of our other observations on the evolution of color terms largely coincide with the thorough analysis presented in Wu 2011.

A.1.13. 'heart': 心 (**səm* > *xīn*). ◊ In the modern language, used primarily as part of the binome 心臟 *xīn-zàng*, literally 'heart-store'. Already in the ancient texts, the word is much more frequently found in abstract meanings ('mind', 'soul', 'conscience', 'intention', etc.) than in the required anatomical meaning; however, there is no evidence whatsoever that Chinese ever knew a different term for the anatomical 'heart'.

A.1.14. 'horn': 角 (**kroːk* > *jiǎo*).

A.1.15. 'I': 我 (**ŋhaːyʔ* > *wǒ*). ◊ For the EOC period, 予 ~ 余 (**dla* > *yú*) must be added as a synonym; the semantic difference between *wǒ* and *yú* is a much debated and still unresolved issue. However, both variants are known already from the Shang period, so there are no arguments in favor of a lexical replacement (merely the elimination of one of the synonyms in the COC period). In COC as well as in certain series of Zhōu epigraphic inscriptions, 我 **ŋhaːyʔ* co-exists with the morphological variant 吾 **ŋha*, but this has no bearing on lexicostatistical calculations, since the root morpheme is obviously the same.

A.1.16. 'kill': 殺 (**sraːt* > *shā*). ◊ There are some signs that in the modern language, the old word 殺 *shā* (or its bisyllabic counterpart 殺死 *shā-sǐ*) is being gradually replaced by the colloquial 打死 *dǎ-sǐ* (lit. 'hit-die'), but 殺 *shā* is still a frequent and stylistically neutral equivalent.

A.1.17. 'know': 知 (**tre* > *zhī*). ◊ Typically used as part of the binome 知道 *zhī-dào* in the modern language. It is useful to note that in the *Línjì lù* dialect this word is in free competition with the synonymous 識 (**tək* > *shì*), whose meaning in COC is closer to 'learn, keep in memory' and in the modern language to 'be acquainted with smbd.'; cf. contexts such as 總識伊來處 «[I] always know the place from which he comes», etc. However, this observation has no impact on the overall statistics for lexical replacements.

A.1.18. 'leaf': 葉 (**lhap* > *yè*). ◊ Extended with the desemanticized suffix 子 in the modern language (葉子 *yè-zi*).

A.1.19. 'many': 多 (**taːy* > *duō*).

A.1.20. 'meat': 肉 (**nhuk* > *ròu*).

A.1.21. 'moon': 月 (**ŋot* > *yuè*). ◊ Typically used as part of the binome 月亮 *yuè-liàng* (lit. 'moon-shine') in the modern language.

A.1.22. 'mountain': 山 (**sraːn* > *shān*).

A.1.23. 'name': 名 (**mheŋ* > *míng*). ◊ Typically used as part of the binome 名字 *míng-zì* (lit. 'name-cognomen') in the modern language.

A.1.24. 'new': 新 (**sin* > *xīn*).

A.1.25. 'night': 夜 (**lias* > *yè*).

A.1.26. 'nose': 鼻 (**bhits* > *bí*). ◊ Extended with the desemanticized suffix 子 in the modern language (鼻子 *bí-zi*).

A.1.27. 'not': 不 (**pə* > *bù*).

A.1.28. 'one': 一 (**ʔit* > *yī*).

A.1.29. 'person': 人 (**nin* > *rén*).

A.1.30. 'rain': 雨 (**whaʔ* > *yǔ*).

A.1.31. 'see': 見 (**keːns* > *jiàn*).

A.1.32. 'sit': 坐 (**ʒoːyʔ* > *zuò*). ◊ The word is only scarcely attested in EOC, and there may be some doubt as to whether it was really the most common and neutral equivalent for 'sit' during that period; a possible competitor is 居 (**ka* > *jū*, with a possible falling tone variant **ka-s*) 'to stay, dwell, reside', for which some contexts might suggest an earlier semantics of 'sit'. There are, however, no strong arguments for taking 坐 *zuò* out of the lexicostatistical comparison; at best, 坐 *zuò* and 居 *jū* could be thought of as synonyms (for the EOC stage only).

A.1.33. 'small': 小 (**sewʔ* > *xiǎo*). ◊ Several more specific adjectives denoting minuscule size are found in the texts (e.g. 細 **seːs* > *xì*, 微 **məy* > *wéi*), but they are statistically infrequent and

never feature in the standard antonymous pair 大 *dà* 'big' vs. 小 *xiǎo* 'small', for which there are multiple examples in the *Shījīng*.

A.1.34. 'stone': 石 (**diak* > *shí*). ◊ Usually extended with the desemanticized suffix 頭 in the modern language (石頭 *shí-tou*).

A.1.35. 'swim': 游 (**lu* > *yóu*). ◊ In the *Línjì lù*, only attested in application to fish (遊魚何得迷 'how did the fish that swim lose their way?'), but no evidence for any different verb denoting the corresponding human activity. In the modern language, mainly used as part of the binome 游泳 *yóu-yǒng*, where 泳 *yǒng* (attested already in the *Shījīng*) seems to be the original equivalent for 'to wade (in water)'.

A.1.36. 'tail': 尾 (**məyʔ* > *wěi*). ◊ Extended with the desemanticized component 巴 *bā* (etymologically = 把 *bǎ* 'handle') in the modern language (尾巴 *wěi-bā*).

A.1.37. 'thou': 汝 (**nhaʔ* > *rǔ*) ~ 爾 (**nheyʔ* > *ěr*). ◊ Both of these variants (freely interchangeable in some texts, dialectally or syntactically conditioned in others), as well as the modern variant 你 *nǐ*, clearly go back to the same root; alternations in the coda sometimes reflect archaic morphology and sometimes irregular dialectal developments, understandable for such high frequency usage forms as personal pronouns. No lexical replacements identified.

A.1.38. 'tongue': 舌 (**lat* > *shé*). ◊ Typically used as part of the binome 舌頭 *shé-tou* (with the same desemanticized suffix as in 'stone' q.v.) not only in the modern language, but already in MC: both the short variant *shé* and the disyllabic form are encountered in the *Línjì lù* as free variants.

A.1.39. 'warm (hot)': 熱 (**ŋet* > *rè*). ◊ For this entry, we choose 'hot' (= 'exceeding tolerable temperature') rather than 'warm', as allowable in the GLD. Unlike 'warm' (OC 溫 **ʔūn* > *wén*; modern 暖 *nuǎn*), 'hot' is quite stable throughout all four stages of Chinese.

A.1.40. 'water': 水 (**tuyʔ* > *shuǐ*).

A.1.41. 'we': 我 (**ŋhaːyʔ* > *wǒ*). ◊ In EOC and COC, sg. 'I' and pl. 'we' were usually not distinguished from each other. From Hàn times on, the differentiation, when necessary, is performed by desemanticized quasi-suffixal morphemes (我公 *wǒ-gōng*, 我等 *wǒ-děng*, 我們 *wǒ-men* etc.) without any replacements for the root morpheme.

A.1.42. 'white': 白 (**braːk* > *bái*).

A.1.43. 'who': 誰 (**duy* > *shuí*). ◊ The morphological derivate 孰 **du-k* (> *shú*), originally 'which one /out of several/?', sometimes replaces the original 誰 *shuí* in some dialects of late OC, but this has no bearing on the overall statistics.

A.1.44. 'woman': 女 (**nraʔ* > *nǚ*). ◊ Used by itself or within the binome 女人 *nǚ-rén* (lit. 'woman-person') in the modern language.

A.1.45. 'yellow': 黃 (**ghʷaːŋ* > *huáng*).

A.2. Items not attested in the *Línjì lù* dialect of MC, but well attested at the three other stages.

A.2.1. 'bird': 鳥 (**tiːwʔ* > *niǎo*). ◊ Initial *n-* in the Běijīng dialect is irregular, but the word is still clearly cognate with its OC predecessors. Should be distinguished from OC 禽 **ghəm* 'game-bird', used mainly in hunting contexts.

A.2.2. 'fat': 脂 (**kiy* > *zhī*). ◊ In the modern language, mainly used as part of the binome 脂肪 *zhī-fáng* (already attested in texts going back to the Jìn dynasty, 3rd–5th centuries A.D.). For both stages of OC, an additional synonym is the word 膏 **kāw* (> *gāo*); semantic difference between **kiy* and **kāw* is impossible to reliably determine based on the available text corpus (in the *Shuōwén jiězì* **kiy* is explained as 'fat of horned cattle' and **kāw* as 'fat of hornless cattle' — an explanation not explicitly confirmed by textual usage, but showing that the two words must have been very close). However, 脂 **kiy* is well attested already in the *Shījīng*, and the existence of an additional synonym is not a reason for postulating a lexical replacement.

A.2.3. 'feather': 羽 (**w/r/aʔ > yǔ*). ◊ In the modern language, normally used as part of the binome 羽毛 *yǔ-máo*, lit. 'feather-hair'.

A.2.4. 'fly /v./': 飛 (**pəy > fēi*).

A.2.5. 'long': 長 (**draŋ > cháng*).

A.2.6. 'round': 圓 (written simply as 員 in the earlier texts; **wran > yuán*). ◊ Attestation in the adjectival meaning in EOC and early COC is extremely scarce and dubious, but verbal ('to be around') and nominal ('circumference') meanings are attested (Schuessler 1987: 791), and there are no other serious candidates for the expression of the adjectival meaning in those periods.

A.2.7. 'sand': 沙 (**sraːy > shā*).

A.2.8. 'seed': 種 (**toŋʔ > zhǒng*). ◊ Extended with the desemanticized suffix 子 in the modern language (種子 *zhǒng-zi*).

A.2.9. 'skin': 膚 (**pra > fū*). ◊ In the modern language, used only as part of the binome 皮膚 *pí-fū*, where 皮 (**bhay > pí*) is also a very old word, encountered much more frequently than **pra* already in EOC (Schuessler 1987: 169, 457); however, its EOC attestations are completely restricted to the notion of 'animal skin', 'fur', 'hide', transparently separating it from the required Swadesh meaning of 'human skin'. The first references to **bhay* as 'human skin' seem to appear no earlier than in Hàn-era texts, and even then mostly as part of the already attested binome 皮膚 *pí-fū* (co-existing with simple *fū*).

A.2.10. 'star': 星 (**sheːŋ > xīng*).

A.3. Items not attested (properly) in EOC, but stable throughout all other periods.

A.3.1. 'ashes': COC 灰 (**smaːy > huī*). ◊ Not attested at all in EOC (nor in the *Línjì lù*, for that matter), but this is the only word with the basic meaning 'ashes' throughout the entire known history of Chinese. Even the graphic shape of the character ('hand' + 'fire') suggests an archaic origin, despite not being attested in epigraphic monuments.

A.3.2. '/tree/-bark': COC 皮 (**bhay > pí*). ◊ It seems that the basic root for 'tree-bark' has always been the same as the root for '/animal/ skin, hide' in general (see A.2.9), although specific instances of 'bark' are lacking in both EOC and the *Línjì lù*. In the modern language, the default equivalent is rather the binomial 樹皮 *shù-pí*, where 樹 *shù* = 'tree'; this does not count as a replacement.

A.3.3. 'bone': COC 骨 (**kuːt > gǔ*). ◊ Strangely enough, the word 'bone' is not at all attested in EOC; however, the graphic shape of the character looks archaic, and there is no specific reason to suggest that the EOC equivalent may have been different. In the modern language the word is usually extended with the desemanticized suffix 頭 (骨頭 *gǔ-tou*).

A.3.4. 'knee': COC 膝 (**sit > xī*). ◊ A somewhat problematic entry; the word 'knee' is not really attested in Chinese until texts typically dated to around the 3rd — 1st cent. BC (*Xún-zǐ*, etc.), nor is it encountered in the *Línjì lù*. Again, however, nothing indicates the existence of any other word in this meaning throughout all the stages of non-dialectal Chinese. In the modern language, the default equivalent is the binome 膝蓋 *xī-gài*, lit. 'knee-cover', that does not count as a replacement.

A.3.5. 'liver': COC 肝 (**kaːn > gān*). ◊ Well attested in COC (though not in early Confucian texts) and MC, but not found in EOC. No indication of any possible alternate equivalents throughout any of the stages of written Chinese.

A.3.6. 'louse': COC 蝨 (**srit > shī*). ◊ Attested in COC (though not in early Confucian texts), but not known in EOC or in the *Línjì lù*. Extended with the desemanticized suffix 子 in the modern language (蝨子 *shī-zi*). The word has a solid Sino-Tibetan etymology (= Tibetan *śig*, Lushai *hrik* 'louse' etc.), indirectly confirming that the word has been super-stable from the beginning.

B. Medium-stable items (31 words)

B.1. Replacements from EOC to COC.

B.1.1. 'breast (= chest)': EOC 膺 (**ʔrəŋ* > *yīng*) → COC 胸 (**sŋoŋ* > *xiōng*). ◊ The latter word is quite clearly the main equivalent for 'male chest' in both COC and the modern language, and is encountered once in the *Línjì lù* in the bound expression 指胸 *zhǐ-xiōng* 'to point at one's breast', which makes it at least a plausible candidate for the same meaning in MC. Conversely, the word is not encountered in any EOC texts, where the only known possible equivalent is 膺 **ʔrəŋ* (although it is largely used in bound expressions and figurative meanings as well). This is sufficient evidence to at least suspect a lexical replacement.

B.1.2. 'man': EOC 夫 (**pa* > *fū*) → COC 男 (**nəːm* > *nán*). ◊ A debatable choice. The assumed replacement **nəːm* is actually well attested already in EOC (Schuessler 1987: 436). However, throughout that period it is encountered infrequently, most often to denote a specific feudal title ('*nán*' = 'baron'); more basic usage is generally confined to the noun phrase 男子 **nəːm-cəʔ* '(male) son', used to specify the gender of the descendant (and thus opposed to 女子 **nraʔ-cəʔ* '(female) daughter'. Schuessler adds several epigraphic examples in which *nəːm* means 'male descendant, son' all by itself and may thus be an abbreviation of **nəːm-cəʔ* (e. g. 我後男 **ŋ^{h}aːyʔ g^{h}oːʔ nəːm* 'my (future) male descendants' [1381 Xuan], etc.). On the other hand, EOC **pa* is statistically far more frequent, and in most contexts, applied to human beings that are male by default (soldiers, farmers, etc.) or expressly meaning 'husband'. It is interesting that in the sole known early literary context in which we encounter the noun phrase 夫人 **pa-nin* [*Shàngshū* 42, 9], it clearly refers to 'man' or 'men', whereas already in COC the term **pa-nin* is more commonly used to denote the wife, i. e. 'man's person', rather than 'man-person'. As for the use of **pa* itself in the COC period, most texts clearly show that it is employed in a «socially marked» manner, either in the derived meaning 'teacher, master' (usually within the compound 夫子 **pa-cəʔ*), or in the meaning 'husband' (often within the antonymous pair 夫婦 **pa-bəʔ* 'husband(s) and wife (wives)'). All of this speaks in favor of a gradual transition from **pa* to **nəːm*, with **pa* still functioning as the main word for 'male person' in Early Zhōu.

B.1.3. 'road': EOC 道 (**lhuːʔ* > *dào*) → COC 路 (**raːks* > *lù*). ◊ In EOC, **lhuːʔ* is the most statistically frequent word denoting the idea of 'road' without any further connotations. It also serves as the basis for the derived verb 導 **l^{h}uː-s* 'to lead, conduct (along the way)' (Schuessler 1987: 116). The word 路 **raːks* 'road' (Schuessler 1987: 395), in comparison, is encountered only in a tiny handful of contexts, most often, within the noun phrase 路車 **raːks kla* 'grand chariot', where it is not even certain that the *raːks* in question represents the same 'word'. It is likely that the gradual replacement of **l^{h}uːʔ* with *raːks* did not really start until COC, possibly caused by the expanding polysemy of the former ('road / way / manner / habit / Tao', etc.).

In COC, the simple word 道 **l^{h}uːʔ* is very rarely employed to denote a physical 'road' by itself — most of the time, it only appears within the compound form 道路 **l^{h}uːʔ-raːks*. On the other hand, 路 **raːks* is very common as 'road' on its own, quite unlike its functions in the EOC period. Likewise, in the modern language the basic equivalent for 'road' is either the bisyllabic 道路 *dào-lù* or the monosyllabic 路 *lù*, but never the monosyllabic 道 *dào*. This fairly transparent shift in usage may count as a lexical replacement, with the original **l^{h}uːʔ* ceding its basic functions to **raːks*.

B.1.4. 'root': EOC 本 (**pəːrʔ* > *běn*) → COC 根 (**kəːn* > *gēn*). ◊ Although the absolute majority of contexts in which **pəːrʔ* is encountered in EOC are metaphorical ('root' as 'foundation', etc.), at least one context [*Shījīng* 255, 8] clearly refers to *pəːrʔ* as 'tree root', opposed to 枝 **ke* 'branches' and **lhap* 葉 'leaves'. The simple pictographic nature of the character also hints at the original semantics of 'tree root'. No other words with this meaning are found in EOC. By contrast, it cannot be doubted that by the end of COC the word 根 **kəːn* had completely replaced

the earlier **pəːrʔ* in the basic meaning 'root (of trees and other plants)', with **pəːrʔ* preserved in a wide range of figurative meanings ('root' as 'origin', 'foundation', 'essentials', etc.). In the *Shuōwén jiězì*, for instance, all of the references to roots of plants always comprise **kəːn*, whereas **pəːrʔ* is reserved for the more abstract meaning 'foundation'.

The difficult problem is to determine the approximate period during which the replacement actually took place. Early Confucian texts offer little help in this matter, since the word 'root' is only encountered in them in figurative meanings ('origin', most of the time), thus, only **pəːrʔ* is attested, but none of the attestations are diagnostic. Cf., however, a diagnostic context in the *Inner Chapters* of *Zhuāngzǐ* [1, 4, 6], a document of comparable antiquity: 仰而視其細枝... 俯而見其大根 *yǎng ér shì qí xì zhī... fǔ ér jiàn qí dà gēn* «he looked up and saw its (the tree's) thin branches... he looked down and saw its big roots». In light of all available evidence, we fill the COC slot with **kəːn*. In the modern language, the situation persists (although the root 根 *gēn* is typically used in binomial constructions, such as *shù-gēn* 樹根 'tree-root', etc.).

B.2. Replacements from COC to MC.

B.2.1. 'belly': OC 腹 (**puk* > *fù*) → MC 肚 ([*dó*] > *dǔ*). ◊ The new word for 'belly' is attested already in the *Línjì lù*: *lǜ niú dǔ lǐ shēng* 驢牛肚里生 '/you/ will be born in the belly of a donkey or a cow'. The new word persists in the modern language, albeit usually extended with the desemanticized suffix 子 (肚子 *dǔ-zi*).

B.2.2. 'burn (tr.)': OC 焚 (**bən* > *fén*) → MC 燒 (**sŋew* > *shāo*). ◊ In EOC, **bən* is the main word for 'burn' and **sŋew* is not attested at all. The latter appears in COC and gradually replaces the former as the most neutral equivalent for the concept: of note may be the statistical observation that in the *Zuǒzhuàn* (5th century BC) we observe 42 cases of **bən* vs. no cases at all of **sŋew*, but in the *Shǐjì* (1st century BC) we already see just 17 cases of **bən* vs. 58 cases of **sŋew* (sporadically, the compound form 焚燒 **bən-sŋew* is also observed). In the *Línjì lù*, the equivalent is either the compound form (e.g. *fén-shāo jīng xiàng* 焚燒經像 «to burn writings and images») or the simple 燒 *shāo* (*bèi huǒ lái shāo* 被火來燒 «you will be burned by fire»); the same situation is typical of the modern language. We may tentatively conclude that **bən* was essentially replaced by **sŋew* around Hàn-era times, i.e. in the interim period between COC and MC.

B.2.3. 'cold': OC 寒 (**gaːn* > *hán*) → MC 冷 (**reːŋʔ* > *lěng*). ◊ The word **reːŋʔ* 'cold' frequently appears in Hàn-era texts, but not in COC, where **gaːn* is still the default equivalent. By MC times, **gaːn* is clearly a bound and archaic form (in the *Línjì lù*, it is only encountered in the idiomatic collocation 寒松 *hán sōng* 'winter pine'), and it remains a bound form in the modern language.

B.2.4. 'eat': OC 食 (**ləkˀ* > *shí*) → MC 喫 (**kheːkˀ* > *chī*). ◊ An early colloquialism attested already in the *Shuōwén jiězì*, *chī* is transparently the neutral equivalent of the meaning 'eat' in the *Línjì lù* (*shí* and *chī* are both attested in the text, but only the latter is regularly encountered in direct speech, e.g. *yī rì chī duō shǎo* 一日喫多少 «how much do they eat per one day?»).

B.2.5. 'eye': OC 目 (**muk* > *mù*) → MC 眼 (**ŋrəːnʔ* > *yǎn*). ◊ The original meaning of the word may have been 'eye-ball' (although already in the *Shuōwén *ŋrəːnʔ* is explained as 目 **muk* 'eye'). In any case, the replacement is quite transparent in the *Línjì lù*, where the old word 目 **muk* is only encountered in bound expressions such as 目前 *mù-qían* 'present', etc.

B.2.6. 'head': OC 首 (**sluʔ* > *shǒu*) → MC 頭 (**dhoː* > *tóu*). ◊ This replacement may have already taken place in Hàn-era time (in the *Shǐ jì*, the word seems to be more frequent than *shǒu*, particularly in direct speech).

B.2.7. 'smoke': OC 熏 (**hun* > *xūn*) → MC 煙 (**ʔiːn* > *yán*). ◊ Available attestations are insufficient to reconstruct a completely reliable picture. The facts so far are as follows: (a) only **hun* is attested in EOC; (b) **ʔiːn* is clearly the main equivalent for 'smoke' in all Hàn-era and later texts; (c) early Confucian texts of the 5th-6th centuries have no occurrences of 'smoke', but the

word is sometimes encountered in texts such as *Mò-zǐ* or *Zhuāng-zǐ*, albeit more often in the verbal ('to smoke out') than nominal meaning. We tentatively assume that the replacement of the original noun has to be dated to a time period around Early Hàn, but new data may overturn this assumption.

B.2.8. 'tree': OC 木 (**mhoːk* > *mù*) → MC 樹 (**dhoʔ* > *shù*). ◊ The nature and reasons for this replacement are quite transparent: it begins as a compound form 樹木 *shù-mù*, lit. 'planted tree' (where 樹 = 豎 **dhoʔ/s/* 'plant vertically'), well attested already in the Hàn period. By late MC, the replacement seems to be complete: in the *Línjì lù*, simple 樹 *shù* is the usual equivalent for 'tree /growing/' (cf. *chéng yī zhū dà shù* 成一株大樹 «he will become a big tree»), while 木 *mù* is restricted to the meaning 'wood /material/'. In the modern language, 'tree' is 樹 *shù* or 樹木 *shù-mù*; 木 *mù* (more frequently, the extended suffixal variant 木頭 *mù-tóu*) is strictly 'wood'.

B.2.9. 'two': OC 二 (**niːs* > *èr*) → MC 兩 (**rhaŋʔ* > *liǎng*). ◊ This only counts as a replacement if we follow the definition of 'two' as an adjectival lexeme, used in conjunction with a quantified noun; since this is the most common function of numerals, such a definition is, however, fully justified. The replacement process is well traceable across ancient texts. The word **rhaŋʔ* is not encountered at all as a numeral in EOC texts; is rigidly restricted to paired objects only throughout COC (兩手 *liǎng shǒu* 'two hands', 兩馬 *liǎng mǎ* 'a pair of horses' etc.); and begins to be freely applied to any objects, paired or not, around Hàn times. In the *Línjì lù* it is clearly the same default equivalent for 'two /of anything/' as it is in the modern language, e.g. 與爾兩文錢 *yǔ ěr liǎng wén qián* 'I give you two coins', etc.

B.2.10. 'go (walk)'[5]: OC 往 (**waŋʔ* > *wǎng*) → MC 去 (**khas* > *qù*). ◊ This replacement is rather tricky and not easily detectable through the corpus, particularly considering the general abundance of verbs denoting directed movement in OC (partial synonyms also include 之 **tə* 'to go, be headed somewhere', 適 **tek* 'to go', etc.). Nevertheless, it can be more or less ascertained that throughout EOC and COC 去 *qù* is almost exclusively used in the meaning 'to /take/ leave', and, even more importantly, that the basic antonymous pair 'come and go' is always rendered as 往來 *wǎng-lái* rather than 往去 *wǎng-qù*. This situation is completely reversed in the language of the *Línjì lù*, where the usual antonym of 來 *lái* is always 去 *qù* rather than 往 *wǎng*, and remains as such in the modern language.

B.2.11. 'what': OC 何 (**ghaːy* > *hé*) → MC 什麼 ([*ʒimmu̯a*] > *shémme*). ◊ While the old inanimate interrogative pronoun still survives in MC as an archaism or as part of some bound expressions, it is clear that already in the *Línjì lù* the default equivalent is the replacement shémme, a colloquialism that arose already in post-Hàn times.

B.3. Replacements from MC to PTH.

B.3.1. 'nail (claw)'[6]: OC 爪 (**cruːʔ* > *zhǎo*) → PTH 指甲 *zhǐ-jiǎ*. ◊ In the *Línjì lù*, the old word 爪 *zhǎo* still seems to be the default equivalent, cf. 髮毛爪齒 *fá-máo zhǎo chǐ* «head hair, body hair, nails, and teeth». The binome 指甲 *zhǐ-jiǎ* (literally 'finger-shell') is first attested in Sòng-era texts (11–12 cent.).

B.3.2. 'give': OC 畀 (**pits* > *bì*) / 予 ~ 與 (**laʔ* > *yǔ*) → PTH 給 *gěi*. ◊ In EOC, **pits* and **laʔ* are basically interchangeable synonyms, cf. two lines in the same *Shījīng* poem (53, 1): 何以畀之 *hé yǐ bì zhī* vs. 何以予之 *hé yǐ yǔ zhī*, both translatable as 'what shall I give him?' Only the latter, however, survives into COC times, where it becomes the sole neutral equivalent for the re-

[5] The meaning 'go' (i.e. the opposite of 'come') is consistently used in the Global Lexicostatistical Database instead of 'walk' (i.e. 'move without a specific direction') in the «classic» Swadesh wordlist, but is still filed alphabetically under 'walk' because of technicalities.

[6] The meaning '(finger)nail' (of human) is consistently used in the Global Lexicostatistical Database instead of 'claw' (animal) in the «classic» Swadesh wordlist, but is still filed alphabetically under 'claw' because of technicalities.

quired meaning and persists into MC. PTH 給 *gěi* is a more recent replacement (a dialectal phonetic development from MC *kip* ← OC **kəp*, originally 'to provide, furnish').

B.3.3. 'green': OC 青 (**sheːŋ* > *qīng*) → PTH 綠 *lǜ*. ◊ Both these words are already attested in EOC and persist all the way to the modern language. Our decision is based primarily on diagnostic contexts, such as the application of these qualifiers to specifically green objects (e.g. 'leaves') and their appearance in lists of the most basic color terms. The latter, in particular, allows to assume that 青 *qīng* was still the basic 'green' as late as MC (cf. in the *Línjì lù*: 把我著底衣, 認青黃赤白 *bǎ wǒ zhuó-di yī, rèn qīng huáng chí bái* «he seizes the clothes that I wear, considers them to be green, yellow, red or white»). In the modern language, however, 青 *qīng* has shifted to denote a darker tinge of green, with 綠 *lǜ* taking its place in the general spectrum.

B.3.4. 'hear': OC 聞 (**mən* > *wén*) → PTH 聽見 *tīng-jiàn*. ◊ The old word is still the default equivalent for 'hear' in the *Línjì lù*; in the modern language, it is only encountered in bound expressions.

B.3.5. 'mouth': OC 口 (**khoːʔ* > *kǒu*) → PTH 嘴 *zuǐ*. ◊ The latter word, originally written simply as 觜, used to mean 'beak'; the shift to 'mouth' is apparently a very recent development that took place sometime in the late Qíng period.

B.3.6. 'red': OC 赤 (**khiak* > *chì*) → PTH 紅 *hóng*. ◊ The latter word is already attested in COC, where it, however, is very rare and most likely denotes some specific shade of red. 赤 *chì* is still the main equivalent for 'red' in the *Línjì lù* (see the example in B.3.2). It is not quite clear at which particular moment the replacement has become complete, but in the modern language 赤 *chì* is no longer in active usage. Other OC words that are typically translated as 'red', e.g. 朱 *zhū*, 彤 *tóng*, etc., are statistically less frequent and more commonly found in conjunction with articles of clothing than natural objects.

B.3.7. 'stand': OC 立 (**rəp* > *lì*) → PTH 站 *zhàn*. ◊ The older meaning of 站 *zhàn* is 'to stop somewhere; to occupy a place' (originally written as 佔). The word gradually replaces the older 立 *lì* in the basic meaning 'to stand' over the Míng-Qíng period.

B.3.8. 'sun': OC 日 (**nit* > *rì*) → PTH 太陽 *tài-yàng*. ◊ The metaphoric term 太陽 *tài-yàng*, lit. 'the extreme Yang', is well attested since at least Hàn times, but only functions as the default term for the celestial body in the modern language.

B.3.9. 'this': OC 此 (**cheyʔ* > *cǐ*) → PTH 這 *zhè*. ◊ There is a certain number of stems that may be used to denote proximal deixis at any given time period in Chinese, but 此 *cǐ* is the one link that ties together all these time periods — with the exception of the modern language, where it is only used in idiomatic bound forms, while the common equivalent for 'this' is the replacement 這 *zhè*. In the *Línjì lù*, both 此 *cǐ* and 這 *zhè* co-exist, but 此 *cǐ* is still far more common and cannot be formally regarded as a literary archaism.

B.3.10. 'tooth': OC 齒 (**thəʔ* > *chǐ*) → PTH 牙 *yá*. ◊ The story here is as follows: (a) in EOC, 齒 *chǐ* = 'teeth /of humans or animals/', 牙 *yá* = '/special/ teeth /of animals only/' (usually tusks, possibly also fangs etc., i.e. protruding teeth; even the graphic shape of the character suggests 'tusks'); (b) in COC, the situation is largely the same, although in a few cases the compound form 齒牙 *chǐ-yá* is also attested; (c) in the *Línjì lù*, the usual equivalent is either bisyllabic 牙齒 *yá-chǐ* or monosyllabic 齒 *chǐ*, but never monosyllabic 牙 *yá*; (d) conversely, in the modern language, the usual equivalent is either bisyllabic 牙齒 *yá-chǐ* or monosyllabic 牙 *yá*, but never monosyllabic 齒 *chǐ*. According to our rules, this indicates a replacement from MC to PTH.

B.4. Unclear due to lack of attestation in MC.

B.4.1. 'dog': OC 犬 (**khʷiːnʔ* > *quǎn*) → PTH 狗 *gǒu*. ◊ Although the word 'dog' is not attested in the *Línjì lù*, it may be reasonably well guessed that 狗 *gǒu* had already become the primary equivalent for the neutral meaning 'dog' in MC, judging by the steady increase in at-

testation since Hàn times, by which period the old 犬 *quǎn* had largely been demoted to the specialized meaning 'hunting dog = hound'. See Starostin 2013 on the possible semantic differentiation between *quǎn* and *gǒu* in COC (where *gǒu* may have originally denoted a special breed of dogs raised for meat).

B.4.2. 'drink': OC 飲 (**ʔəmʔ* > *yǐn*) → PTH 喝 *hē*. ◊ Not attested in the *Línjì lù* at all. The modern equivalent 喝 *hē* is only encountered in texts since the Yuán dynasty (13th — 14th centuries), so it may be assumed that the old word was still in colloquial circulation throughout the MC period.

B.4.3. 'egg': COC 卵 (**rhoːnʔ* > *luǎn*) → PTH 蛋 *dàn*. ◊ The old word is not attested either in EOC (although the pictographic nature of the character may suggest an archaic origin) or in the *Línjì lù*. The new word is a transparent semantic extension of *dàn* 'ball, pill, bullet, any small round object', a word well attested already in OC and usually written as 彈 in its original meaning. The first attestations of the semantic shift come from classic 16th–18th century novels; it may be assumed that the old word *luǎn* was still the basic term in MC[7].

B.4.4. 'full': OC 盈 (**leŋ* > *yíng*) → PTH 滿 *mǎn*. ◊ Not attested in the *Línjì lù*. The original meaning of 滿 *mǎn* was likely 'to fill up, overflow (of water)'; it is not found in the generic meaning 'to fill /anything/' or in the adjectival meaning 'full' in early Confucian texts or in the *Dàodéjīng*, but is already competing with 盈 *yíng* in *Zhuāngzǐ*. In the *Shījì*, 盈 *yíng* is encountered 14 times next to 85 for 滿 *mǎn*, meaning that the replacement was likely complete by the early Hàn period.

Another semantically close morpheme, 充 (**thuŋ* > *chōng*), is first encountered in the *Shījīng* as part of the compound noun 充耳 *chōng-ěr* 'ear stopper'; in COC it is usually applied to the process of filling up storage units (granaries, etc.) and also used in various figurative meanings. The bisyllabic compound 充滿 *chōng-mǎn* is well attested already in Early Hàn times and has persisted all the way up to modern times; nevertheless, 充 *chōng* almost always behaves as a secondary morpheme in this formation, and while it is hard to precisely state the semantic difference between *chōng* and *mǎn* in the COC period (it may have been 'to fill up with hard substances' vs. 'to fill up with liquid substances', as one of the possibilities), including it in our calculations as a secondary synonym or excluding it altogether will have no effect on the overall calculations.

B.4.5. 'neck': OC 領 (**rheŋʔ* > *lǐng*) → PTH 脖子 *bó-zi*. ◊ Not attested in the *Línjì lù*. Modern *bó-zi* is a very late word, not attested earlier than the Yuán dynasty (13th–14th centuries). In addition, a very frequent equivalent for 'neck' in early Hàn texts is OC 項 **groːŋʔ* (> *xiàng*), whereas 領 is more frequently used in the meaning 'collar' by that time. It cannot, however, be confirmed at this time that 項 *xiàng* continued to be the main term for 'neck' throughout MC. Another occasional synonym in COC is 脰 (**dhoːs* > *dòu)*, always translated as 'neck'; in about 90% of its occurrences in texts, it is used as the object of 'breaking' or 'cutting', implying immediate death, so it is possible that a more exact meaning is something like 'neck vertebra'. In any

[7] It is suggested in Baxter, Sagart 2014: 324 that a more archaic equivalent for 'egg' may be a root **tʰu[n]* (= **tʰun* or **tʰur*), not attested in any written Chinese texts but functioning as a vulgar equivalent for 'egg' and/or 'testicles' in some Southern dialects (Cantonese *tʃʰœn*¹, Hakka *tʃʰun*¹); its antiquity is allegedly corroborated by semantically and phonetically perfect Tibeto-Burman parallels. Regardless of whether this hypothesis is correct, it could only be taken into consideration in this paper if we were to assert that this **tʰu[n]*, not 卵, had the basic meaning 'egg' in EOC, and that somehow Cantonese and Hakka had managed to inherit it, completely bypassing the COC and MC stages. Since the first part of this statement has no confirmation in written evidence and the second is almost impossible to believe, at best we could hypothesize that **tʰu[n]* may have existed in EOC and COC side-by-side with 卵 as a «vulgar» synonym, managing to survive into Cantonese and Hakka; but this hypothesis would have no bearing on our lexicostatistics, which requires that only the stylistically neutral equivalents be taken into consideration.

case, it is a statistically infrequent (no more than a couple dozen entries in the entire COC + Hàn corpus, next to hundreds for 領 **rheŋʔ* and 項 *xiàng*) and contextually bound word.

B.4.6. 'that': OC 彼 (**payʔ* > *bǐ*) → PTH 那 *nà* ~ *nèi*. ◊ Not attested in the *Línjì lù*, although apparently certain other texts in the *yǔlù* genre already show 那 *nà* as the basic adjectival stem denoting objects that are far away, while 彼 *bǐ* is more frequently restricted to adverbial functions ('there', 'in that place'). On the other hand, cf. B.3.9 'this' where it can be seen that both the old and the new pronoun still co-exist in the *Línjì lù* dialect as synonyms; it cannot be excluded that the same situation was symmetrically relevant for the distal deixis pronouns.

C. Unstable items (5 words)

C.1. EOC → COC, COC → MC.

C.1.1. 'bite': EOC 啑 (**di:t* > *dié*) → COC 噬 (**dats* > *shì*) → MC 咬 (**ŋhra:wʔ* > *yǎo*). ◊ The double replacement is quite uncertain[8]: so far, the only unambiguous EOC context with the verb 'to bite' is a passage in the earliest layer of the *Yìjīng*: 履虎尾, 不啑人 *lǚ hǔ wěi, bù dié rén* «if one steps on a tiger's tail, he does not bite». The situation in COC is also far from clear: statistically and contextually, there is some serious competition for 噬 **dats* on the part of 齧 (**ŋhe:t* > *nié*), also encountered several times (*Zhuāng-zǐ*; *Guǎn-zǐ*) in the meaning 'to bite' (or perhaps 'to gnaw?') as applied to dogs. The distinction between **dats* and *ŋhe:t* may have originally been dialectal (e. g. «Northern» vs. «Southern»), but it becomes seriously blurred in Hàn times (thus, both terms are interchangeable in the *Huáinán-zǐ*). Since MC, however, 咬 *yǎo* seems to have largely stabilized as the primary equivalent for this meaning.

C.1.2. (?) 'foot': EOC 趾 (**təʔ* > *zhǐ*) → COC 足 (**cok* > *zú*) → MC 腳 (**kak* > *jiǎo*). ◊ The fact that the 'foot' / 'leg' opposition in the earliest stages of Chinese was lexicalized as 趾 (originally written simply as 之) *zhǐ* 'foot' vs. 足 *zú* 'leg' is suggested, first and foremost, by the early graphical shapes of the characters:  'foot' vs.  'leg'. Textual evidence is ambiguous at best, since both 'feet' and 'legs' are very rarely attested in EOC, but at least one context in the *Shījīng* (麟之趾 *lín zhī zhǐ* 'the feet (= hooves) of the *lín*') indirectly supports this difference. In COC the old word *zhǐ* seems to have shifted its meaning to 'toe', while both 'foot' and 'leg' seem to merge into 足 *zú* for a while — at least until Hàn-era texts, when the differentiation re-emerges with the appearance of a new word for 'foot', 腳 *jiǎo* (not attested in EOC at all).

C.1.3. 'sleep': EOC 寐 (**miys* > *mèi*) → COC 臥 (**ŋho:ys* > *wò*) or COC 寢 (**shimʔ* > *qǐn*) → MC 睡 (**doys* > *shuì*). ◊ In EOC, 寐 **miys* is the most common designation of the static meaning 'sleep'; 寢 **shimʔ* is more rare and better interpreted as the dynamic 'lie down to sleep', or causative 'put to sleep' (antonymous to 興 *xīng* 'rise'). In COC, 寐 **miys* is practically non-existent, whereas 寢 **shimʔ* is sometimes found in unambiguously static contexts (e.g. 宰予晝寢 *zǎi yú zhòu qǐn* «Zai Yu slept during the day» [Lùnyǔ 5, 10]); however, it seems to be competing for the 'sleep' slot with 臥 **ŋho:ys*, a word that can be interpreted as 'to lie' or 'to sleep' depending on the context. By Hàn times, the word 睡 **doys* makes its appearance, and seems to completely eliminate all competition by the beginning of the MC period.

C.2. EOC → COC, MC → PTH.

C.2.1. 'all': EOC 率 (**srut* > *shuài*) or 咸 (**gro:m* > *xián*) → COC 皆 (**kro:y* > *jiē*) → PTH 都 *dōu*. ◊ We equate 'all' with the most commonly used Chinese adverbial adjuncts with the same meaning, typically placed right before the verb. EOC uses a variety of those, making it impos-

[8] Laurent Sagart (p.c.) has suggested the possibility of both **di:t* and **dats* reflecting the same original root, but the vocalism seems to go against this idea; even if this were so, the morphological alternation must have been so ancient that the two forms would hardly feel related in the 1st millennium BC.

sible to choose between **srut* and **grəːm*. In COC, 皆 **krəːy* is unquestionably the most widely used adjunct, although by early Hàn times it begins to compete with the synonymous 悉 (**sit* > *xī*); in the *Línjì lù*, **krəːy* is still encountered either on its own or in conjunction with **sit* (both 悉皆 *xī-jiē* and 皆悉 *jiē-xī* are possible). Curiously, modern 都 *dōu* seems to have already existed in its current meaning at least in Hàn times, but is only very occasionally attested until the modern phase of the language.

C.3. COC → MC, MC → PTH.

C.3.1. 'say': EOC 曰 (**wat* > *yuē*) → MC 云 (**wən* > *yún*) → PTH 說 *shuō*. ◊ We understand 'say' here as the most common verb to introduce direct speech, which makes it easier to single out one particular candidate among a huge variety of verbs denoting various kinds of speech in Chinese. In Old Chinese, this verb has always been 曰 **wat*; in the *Línjì lù*, direct speech is usually introduced by 云 **wən*, a verb already well attested in OC as well but nowhere near as common as **wat* (its functions in various subperiods and dialects are still somewhat unclear). In colloquial PTH, the functions of these words have been completely overtaken by 說 *shuō*, a word originally meaning 'to explain, interpret'.

D. Unusual deviations

These two cases describe interesting situations where one of the two intermediate attested stages features a variant that is deviant of the common form, so that older and newer forms of the language share the same equivalent but the intermediate equivalent is expressed by a different root.

D.1. 'earth': EOC 土 (**thaːʔ*) → PTH *tǔ* vs. MC 地 (*dî*). ◊ The semantic difference between 土 *tǔ* and 地 *dì* 'earth, ground' is often neutralized in both ancient and modern contexts, most obviously so within the compound formation 土地 *tǔ-dì*, well attested already in OC. Nevertheless, whenever the two morphemes are met separately, the former typically refers to 'earth' as substance ('soil' — the required Swadesh meaning) and the latter as surface ('ground', 'territory'). Surprisingly, one glaring exception is the dialect of *Línjì lù*, where it is 地 *dì* rather than 土 *tǔ* that commonly functions as a substance term, cf.: 被地水火風 *bèi dì shuǐ huǒ fēng* «suffer earth, water, fire, and wind» (the elements), etc., whereas the word 土 *tǔ* is almost always encountered only within the compound form 國土 *guó-tǔ* «territory (of state)». It is possible that this usage reflects a genuine case of lexical replacement in the respective dialect, though a specific peculiarity of the literary language is not excluded either.

D.2. 'good': EOC 好 (**huːʔ*) → PTH *hǎo* vs. COC 善 *danʔ* (→ PTH *shàn*). ◊ Curiously, the character 好 throughout most of the Classical Chinese period is most often employed to transcribe the derived verbal stem *huː-h* 'to love' rather than the original adjectival stem *huːʔ* 'good' (as in EOC); the latter cannot by any means pretend to denote the basic qualitative predicate '(to be) good' in any of the early Confucian texts or, in fact, in any of Classical Chinese up at least to the Hàn period. Thus, it is a rare (but not unique) isogloss that places EOC closer to post-Classical language than to the Classical epoch. Other quasi-synonyms have been excluded from comparison, such as 佳 (**krē* > *jiā*) 'beautiful, excellent' (met more rarely and generally in highly expressive contexts), 良 (**raŋ* > *liáng*) 'kind, good-spirited' (usually applied to human or animal nature rather than anything else), etc.

E. Excluded from analysis

E.1. 'lie': This (static) meaning is notoriously hard to separate from the closely related 'lie down, go to sleep' (dynamic) and 'sleep', not only in ancient texts, but in many modern dialec-

tal corpora as well (it is no wonder that it is very frequently omitted from various wordlists published in Chinese linguistic sources). The PTH equivalent is the recent innovation 躺 *tǎng*, of unclear origin; earlier literary sources mostly feature ambiguous data, with such quasi-synonyms as 寢 *qǐn* and 臥 *wò* translatable as 'go to sleep', 'lie down', or 'be sleeping' depending not only on the context, but on the translator's intuition as well. There is no formal ground in this case to speculate on possible lexical replacements in pre-PTH times.

Analysis

Having presented the data in its entirety, we can now proceed to the stage of analysis — a relatively brief one, since our only important task here is to calculate the number of replacements (or, more accurately, discrepancies, since we do not want to assume that each of the four analyzed stages was a direct linguistic descendant of the previous one). As could already be seen from the data, many cases in which such discrepancies were postulated are actually problematic and often derived from indirect evidence, particularly in the case of EOC vs. COC, where the attested corpus does not always allow us to resolve the issue of synonimity to complete satisfaction. For that reason, in the tables below I will discriminate between «certain» and «probable» replacements, where the former are clearly evident from sufficient textual evidence and the latter are based on insufficient and/or circumstantial evidence.

Additionally, in respect to the long transitional period from COC to MC it is useful to log the information on cases where a solid argument may be made for a lexical replacement already evident in Hàn-era literary texts (despite the lack of a separate wordlist for the Hàn period); such cases will be marked with a + sign next to the item in question.

	Certain replacements	Probable replacements
EOC → COC	'all', 'road', 'root', 'sleep'	'bite', 'breast (chest)', 'foot', 'good', 'man'
COC → MC	'belly', 'bite', 'cold', 'dog+', 'eat', 'eye', 'foot+', 'head+', 'say', 'sleep', 'tree', 'two', 'go', 'what'	'burn+', 'earth', 'smoke+', 'full+', 'neck+', 'that'
MC → PTH	'all', 'nail', 'give', 'hear', 'mouth', 'red', 'stand', 'sun', 'say'	'green', 'this', 'tooth', 'drink', 'egg'

Adding up both certain and probable replacements, we thus get the following picture:

1) 9 replacements over the approximately 400–500 year period separating EOC from COC;

2) 20 replacements over the approximately 1,200–1,400 year period separating COC from MC (of these, about a third may have taken place over the approximately 300-200 year period separating COC from Hàn-era Chinese, though this number is not fully confirmed);

3) 14 replacements over the approximately 800–1,000 year period separating MC from PTH;

4) altogether, 43 replacements from EOC to PTH (counting twice for those few items that have been replaced two times — 38 otherwise).

Quite importantly, none of the attested replacements can be reliably attributed to external borrowing; although for some of them (especially those that lack reliable Tibeto-Burman cognates) an original non-Chinese source is quite possible, the majority are first attested in texts with non-Swadesh meanings, so the replacements have to be judged as «internal». According to Sergei Starostin's revised methodology of glottochronological calculations, this means that we should expect the rates of change to be reasonably regular, without any periods of intensive speeding-up due to contact-induced processes of lexical intereference.

The results are not convincingly consistent with the division of the Swadesh wordlist into the less stable and more stable sub-sets as described, e.g., in Starostin 2010: although of all the

listed items, slightly less than half belong to the more stable sub-set ('nail', 'dog', 'drink', 'eat', 'egg', 'eye', 'foot', 'head', 'hear', 'mouth', 'smoke', 'sun', 'tooth', 'tree', 'two', 'what'), the proportion is still close to 50/50 and hardly significant. It does seem interesting that nearly all the reliable and potential replacements from EOC to COC fall into the less stable half of the wordlist, but whether this observation is historically important remains to be seen.

Conclusions

1. Taking Early Old Chinese as the starting point and Modern Chinese as the endpoint, we can claim, based on a mix of direct and indirect evidence from the text corpus (and some dictionary information), that *approximately* 60% of the Swadesh wordlist has been retained over 3,000 years of linguistic evolution. (The rounding-up of the percentage, rather than being an aesthetic concession, should hint at the possibility of errors in data analysis and occasional wrong conclusions based on insufficient data). This figure is not in direct contradiction either with the classic Swadesh formula (t = -ln(0.6) / 0.14 ≈ 3650 years) or with the revised Starostin formula ($t = \sqrt{}$ -ln(0.6) : 0.05×0.6 ≈ 4120 years), though it does obviously fit in better with Swadesh's assessment.

2. The individual replacement rates for the three checkpoints are as follows: ≈ 0.18 for EOC to COC, ≈ 0.14 for COC to MC, ≈ 0.14 for MC to PTH. Other than a slight increase in the first case (which could be explained by different factors, such as incorrect dating, errors in wordlist construction, or a significantly divergent dialectal base for EOC, meaning that the real time difference between it and COC should be higher), the results over different time periods seem to be impressively consistent — *and* in unexpectedly good agreement with Swadesh's classic *lambda* value of 0.14 for 1,000 years (rather than Sergei Starostin's 0.05 over the same period).

3. However, these figures may need slight corrections depending on whether we subscribe to the idea that the selected checkpoints are not necessarily in a straightforward ancestral relationship: for instance, the real time distance between MC and PTH may not be the 800–1,000 years that separate the text of the *Línjì lù* from today's colloquial Mandarin Chinese, but a period of as much as 1,000–1,400 years (to be more confident, one would have to conduct a very thorough and rigorous dialectal study of the text). In other words, observed *lambda* values might be slightly inflated (but only slightly: thinking of MC and PTH as two completely independent developments from COC or EOC is not supported by evidence).

4. If there is any circumstantial evidence for a one-time acceleration period, the best candidate would probably be the transition from COC to Hàn-era texts, where we witness, over a span of no more than 200 years, the replacement of such words as 'head', 'neck', 'foot', 'dog', and others. However, since the main dialect of Hàn-era texts is hardly a direct descendant of the Northern (Lǔ?) dialect that forms the basis for the COC list, it may be argued that at least some of these replacements could have happened earlier and are simply undetected due to lack of textual evidence from that dialect preceding the 3rd century BC (which brings us back to point 3).

5. It is particularly instructive to compare the acquired result with historically similar situations for other written languages, especially those already covered in the Global Lexicostatistical Database (Starostin ed. 2011–2019). Thus, for the Greek language (wordlists compiled and published by Alexei Kassian) we have a wordlist for the Ancient Attic dialect (4th century BC, largely based on the language of Plato), compared with Modern Demotic Greek: the number of lexical replacements is 39 (all of them internal, just like in Chinese), which gives a *lambda* value of ≈ 0.16, completely in line with our results for Chinese (unfortunately, no high quality wordlists for any forms of Byzantine Greek are as of now available in the GLD).

On the other hand, it is also true that comparison with another Indo-European situation, namely, Old Norse vs. Modern Icelandic, shows a different result: only 2 replacements ('eat', 'swim') over the approximately 700-800 years that separate the two stages, resulting in a lambda value of ≈ 0.025 (this result basically just repeats the observations already publicized in the well-known anti-glottochronological paper by Bergsland and Vogt, 1962). But what this shows, in my opinion, is not the simplistic «glottochronology does not work» conclusion that is drawn by many researchers, but rather that different rates of replacement may be triggered by different sociolinguistic situations — indeed, it may be argued that historically, the cases of Greek and Chinese have more in common with each other (large dialectal variety; co-existence of an archaic written language with evolving colloquial norms; active contact with neighboring languages) than either of them with Icelandic. Naturally, a full comparative analysis of these situations will only be possible after a detailed analysis of all the empirical evidence that may be gathered from other written languages across the globe (Indo-European, Semitic, Egyptian, etc.); hopefully, the present study takes a small step in the right direction.

References

Baxter, William H., Laurent Sagart. 2014. *Old Chinese: A New Reconstruction*. Oxford University Press.

Bergsland, Knut, Hans Vogt. 1962. On the Validity of Glottochronology. *Current Anthropology* 3: 115–153.

Dobson, W. A. C. H. 1968. *The Language of the Book of Songs*. University of Toronto Press.

Gurevich, Isabella S. 2001. *Lin-ji lu*. Saint-Petersburg: Peterburgskoje vosotokovedenije.

Hamed, Mahé Ben, Wang Feng. 2006. Stuck in the forest: Trees, networks and Chinese dialects. *Diachronica* 23(1): 29–60.

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

List, Johann-Mattis. 2015. Network perspectives on Chinese dialect history. *Bulletin of Chinese Linguistics* 8: 42–67.

List, Johann-Mattis. 2016. Beyond cognacy: Historical relations between words and their implication for phylogenetic reconstruction. *Journal of Language Evolution* 1(2): 119–136.

Pulleyblank, Edwin G. 1995. *Outline of Classical Chinese Grammar*. Vancouver: UBC Press.

Sawer, Michael. 1969. *Studies in Middle Chinese Grammar: the language of the early yeu luh*. PhD Thesis, Australian National University.

Schuessler, Axel. 1987. *A Dictionary of Early Zhōu Chinese*. Honolulu: University of Hawaii Press.

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

Starostin, George. 2013a. Lexicostatistics as a basis for language classification: increasing the pros, reducing the cons. In: H. Fangerau, H. Geisler, Th. Halling, W. Martin (eds.). *Classification and Evolution in Biology, Linguistics and the History of Science: Concepts — Methods — Visualization*: 125–146. Stuttgart: Franz Steiner Verlag.

Starostin, George. 2013b. K probleme dvux sobak v klassicheskom kitajskom jazyke: canis comestibilis vs. canis venaticus? In: N. P. Grintser et al. (eds.). *Institutionis Conditori: Ilje Sergeevichu Smirnovu. Orientalia et Classica, vol. L*: 253–267. Moscow: RSUH Publishers.

Starostin, George (ed.) 2011–2019. *The Global Lexicostatistical Database*. Moscow: Russian State University for the Humanities, & Santa Fe: Santa Fe Institute. Available online at: http://starling.rinet.ru/new100.

Starostin, Sergei A. 1989. *Rekonstruktsiia drevnekitaiskoi fonologicheskoi sistemy* [Reconstruction of the Old Chinese phonological system]. Moscow: Nauka. (In Russian).

Starostin, Sergei. 2000. Comparative-historical linguistics and lexicostatistics. In: Colin Renfrew, April McMahon, Larry Trask (eds.). *Time Depth in Historical Linguistics*: 223–259. McDonald Institute for Archaeological Research, Oxford Publishing Press.

Sturgeon, Donald (ed.). 2019. *Chinese Text Project*. Available online at: https://ctext.org.

Swadesh, Morris. 1952. Lexico-Statistic Dating of Prehistoric Ethnic Contacts: With Special Reference to North American Indians and Eskimos. *Proceedings of the American Philosophical Society* 96(4): 452–463.

Wu, Jianshe. 2011. The evolution of basic color terms in Chinese. *Journal of Chinese linguistics* 39(1): 76–122.

Г. С. Старостин. Китайская базисная лексика в диахронической перспективе и ее значимость для лексикостатистики и глоттохронологии

В статье сравниваются относительные скорости замены базисной лексики (представленной стандартным 100-словным списком Сводеша) на протяжении истории развития китайского языка, от раннедревнекитайского (представленного такими текстами, как *Книга песен*) к классическому древнекитайскому, позднему среднекитайскому (представленному языком памятника *Линьцзи лу*) и современному китайскому. В первой части статьи последовательно излагается методология составления списков; вторая посвящена детальному обсуждению всех обнаруженных лексических замен. В заключительной части показано, что в среднем скорость распада списка от одного периода к другому меняется незначительно, и что в целом результаты согласуются с классической «константой Сводеша» (0.14 замен за тысячу лет); более того, обнаруживается корреляция и с некоторыми другими аналогичными ситуациями, например, с историей греческого языка, хотя в отдельных случаях (исландский) такой корреляции не наблюдается. Можно надеяться, что дальнейшие исследования такого рода по лексической эволюции языков с длительной письменной историей позволят поместить полученные результаты в более широкий и значимый контекст.

Ключевые слова: история китайского языка, древнекитайский язык, среднекитайский язык, лексикостатистика, глоттохронология, базисная лексика.