



HAL
open science

NO LIMITS TO BORROWING: THE CASE OF BAI AND CHINESE

Yeon-Ju Lee, Laurent Sagart

► **To cite this version:**

Yeon-Ju Lee, Laurent Sagart. NO LIMITS TO BORROWING: THE CASE OF BAI AND CHINESE. *Diachronica*, 2008, 25 (3), pp.357-385. <10.1075/dia.25.3.03yeo>. <hal-00781005>

HAL Id: hal-00781005

<https://hal.science/hal-00781005v1>

Submitted on 25 Jan 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

NO LIMITS TO BORROWING: THE CASE OF BAI AND CHINESE

Lee Yeon-Ju (Kangwon National University, Chuncheon, Korea)

and

Laurent Sagart (Centre de Recherches Linguistiques sur l'Asie Orientale, Paris, France)

keywords: Sino-Tibetan, Chinese, Bai, contact, stratification, subgrouping, numerals

Abstract

Based on the large amount of Chinese-related basic vocabulary in Bai, scholars like Benedict, Starostin and Zhengzhang have claimed a special phylogenetic proximity between Bai and Chinese. In this paper we show that the Chinese vocabulary in Bai is stratified, forming successive layers of borrowings. We identify three such layers, describing the sound correspondences which characterize each of them: two Mandarin layers, one local, one regional for modern words; and an early Chinese layer, acquired during a long and complex period of intimate contact between Bai and Chinese, beginning in Han times and terminating in Late Tang, altogether a millennium or so. This last layer is subdivided into several sub-layers. The remaining part of the vocabulary forms the Bai indigenous layer, whose affiliation is clearly Sino-Tibetan, without having any particular proximity to Chinese. In particular, the numerals '1' and '2' have etymological connections among non-Chinese Sino-Tibetan languages such as Jingpo, Sulung and Tangut. The numerals above "2" are Chinese loanwords and even the numerals "1" and "2" have less colloquial variants of Chinese origin. Bai is of interest to comparative linguistics for the extraordinary amount of basic vocabulary it has borrowed from Chinese, all of it during the early period: 47% of the 100-Swadesh list.

Résumé

S'appuyant sur l'abondant vocabulaire de base commun au bai et au chinois, des auteurs comme Benedict, Starostin et Zhengzhang ont affirmé qu'il existe une proximité phylogénétique particulière entre ces deux langues. Nous montrons ici que le vocabulaire du bai est stratifié, et qu'il faut y distinguer trois couches chronologiques, dont nous décrivons les correspondances phonétiques avec le chinois. Les deux premières sont formées d'emprunts récents à deux variétés distinctes de mandarin du sud-ouest, l'une locale, l'autre régionale ; la troisième est une couche d'emprunts anciens, acquis au cours d'une longue période de contact d'environ un millénaire, de l'époque Han à la fin des Tang. Cette couche est elle-même subdivisée en plusieurs sous-couches. Le reste du vocabulaire forme la couche indigène : elle est d'affiliation clairement sino-tibétaine, mais sans proximité

particulière avec le chinois. Notamment, les nombres "1" et "2" peuvent être comparés aux nombres correspondants en Jingpo, Sulung et en Tangoute. Les nombres au-dessus de "2" ont été empruntés au chinois, et même les nombres "1" et "2" ont des variantes littéraires d'origine chinoise. L'intérêt du Bai pour la linguistique comparative tient au nombre exceptionnel de mots du vocabulaire de base empruntés au chinois, depuis les Han jusqu'aux Tang : 47% de la liste de cent mots de Swadesh.

Zusammenfassung

Aufgrund des umfangreichen, auf der chinesischen Sprache basierenden Grundvokabulars in Bai haben Gelehrte wie Benedict, Starostin und Zhengzhang eine phylogenetische Verwandtschaft zwischen Bai und Chinesisch erkennen wollen. In dieser Arbeit wird gezeigt, dass der chinesische Wortschatz in Bai stratifiziert ist, d. h. er bildet Schichten von Lehnwörtern. Wir haben drei solcher Schichten herausgearbeitet und beschreiben die Lautgesetze, die für jede Schicht charakteristisch sind: zwei Schichten des Mandarin, eine lokale und eine regionale für moderne Wörter, sowie eine ältere chinesische Schicht, die im Laufe einer langen Periode enger Interaktionen zwischen Bai und Chinesisch entstanden ist. Diese Periode begann in der Han-Zeit und endete in der späten Tang-Zeit – das ergibt also eine Zeitspanne von ca. 1000 Jahren. Diese letzte Schicht ist in mehrere Unterschichten aufgeteilt. Der verbleibende Teil des Vokabulars bildet die indigene Baischicht, welche klar dem Sino-Tibetanischen zuzuordnen ist, jedoch ohne irgend eine besondere Verwandtschaft zum Chinesischen aufzuweisen. Insbesondere Nummer 1 und 2 besitzen etymologische Verbindungen zu nichtchinesischen Sprachen wie Jingpo, Sulung oder Tangut. Für vergleichende Sprachwissenschaft ist Bai interessant wegen des außergewöhnlich großen, gänzlich während der frühen Periode aus dem Chinesischen übernommenen Grundvokabulars: 47% der 100-Worte-Swadesh-Liste.

Introduction

Bai is a Sino-Tibetan language spoken in Yunnan. Its affiliation within Sino-Tibetan is disputed. The classical view (Li 1937; Dell 1981; Zhao 1982; Lee & Sagart 1998) is that Bai is a Tibeto-Burman language that has received very strong Chinese influence, especially in its vocabulary. Other scholars (Benedict 1982, Starostin 1995b, Zhengzhang 1999), noting the very large amount of basic vocabulary shared by Chinese and Bai, regard Bai as most closely related with Chinese within Sino-Tibetan (Benedict) or even as an early dialect of

Chinese (Starostin). For discussions of the history of the Bai language see Bradley (1979), Wiersma (1990, 2003).

In many familiar cases of lexical borrowing, loans from a donor language form a single, well identifiable layer, within the recipient. However, when Chinese is the donor to a language with which it has a long history of contact, the situation is different. Chinese culture has experienced over time a succession of periods of expansion and contraction: as shown by Norman (1979), during periods of Chinese cultural expansion ("waves of sinicity"), important numbers of loanwords are issued to languages in contact –including Chinese "dialects". In any given language, then, Chinese loanwords are stratified, forming several distinct chronological layers, each with its specific correspondence rules.

This paper is a study of the stratification of the vocabulary of one dialect of Bai, spoken in Jianchuan 劍川. The data are drawn from Huang et al. (1992, language #48), a lexical atlas of Sino-Tibetan languages in China. The groundwork for this study was conducted in 1997-1998 by the authors in Geneva and Paris. A preliminary report was made at a conference in Lund in 1998 (Lee & Sagart 1998).¹ Our aim in that paper was, first, to clarify the stratification of Chinese loanwords to Bai, and second, to reassess the question of the affiliation of Bai taking into account the stratification of its lexicon. Our tool for analyzing this stratification was the coherence principle (first explicitly formulated in Sagart & Xu 2001):

"the initial, rhyme and tone correspondences on a borrowed syllabic morpheme obey the same set of correspondences"

(Sagart & Xu 2001:15)

This principle states that in borrowed syllabic morphemes, all correspondences come from a single layer or stratum. To those who regard sound change as essentially regular, this is close to a truism. However, especially in China, Hong Kong and Taiwan, some scholars associated with W. S.-Y. Wang's theory of Lexical Diffusion claim that borrowed sounds will compete with other borrowed sounds within the lexicon of the recipient language, in effect creating situations where the correspondences on a borrowed syllabic morpheme come from different layers.

¹ See <http://halshs.archives-ouvertes.fr/halshs-00250200/fr/> .

In the course of analyzing the data, one of us (Lee Yeon-ju) discovered that the principle extended to disyllables too. This led to the formulation of the extended principle of coherence in Sagart and Xu (2001).

"The initial, rhyme and tone correspondences on all syllables of one borrowed polysyllabic morpheme obey the same set of correspondences, provided the morpheme is semantically noncompositional"

(Sagart & Xu 2001:16)

This means that in borrowed disyllabic words, both syllables belong to the same layer. Semantic noncompositionality was selected as a protection against hybrid forms, i.e. compound words with morphemes drawn from different borrowing layers. Such hybrid forms were not borrowed as units, but were assembled within Bai from morphemes borrowed at different periods. There are many such examples in Bai. An example is 'fist' *si33 tchuẽ55* where the first syllable, from Chinese 手 'hand', MC *syuwX, belongs to our early layer, while the second, from Chinese 拳 'fist' (Mandarin *tɕ'yan35*), is a Mandarin loanword. However, the great majority of borrowed disyllables in Bai are from a disyllabic Chinese word and the extended principle of coherence applies. This principle is a very powerful tool for working out the stratification of Chinese loanwords in a language like Bai, because each borrowed disyllable presents us simultaneously with two correspondences of syllable onsets, two correspondences of rimes and two correspondences of tones, all from the same layer. Tones are particularly useful for the purpose of discriminating between layers. In most cases, two-tone combinations are distinctive enough to permit correct layer assignment, even in the case of varieties of southwestern Mandarin which are mutually perfectly intelligible, as we show below.

The present study is our final report on our work. It is based on our analysis of Bai disyllables, and it relies principally on the extended principle of coherence.

Bai syllables are C(G₁)V(G₂) (where 'G' stands for a nonsyllabic high vowel) in structure. Vowels are either tense or lax, and either oral or nasal. For an outline of Jianchuan Bai phonology, see Huang et al. (1992: 675-676, hereafter TBL; Xu and Zhao 1984 for a slightly different account). Here we limit ourselves to reproducing tables of initials consonants and tones from Huang et al.-

Bai initial consonants

P	ts	t	tɕ	k
Ph	tsh	th	tɕh	kh
M		n	ɲ	ŋ
F	s		ɕ	x
V	z			ʎ
		l	j	

Note: p t k ts tɕ are voiced when occurring with tones 33 and 21

Bai tones

55 42 35 33 21²

Notes:

- Bai tone 35 occurs only in recent Chinese loanwords in the Entering tone. See Table 3 for examples.
- 42 has 'mixed creaky phonation' (聲門混合擠擦音), 21 has breathiness (氣化現象)
- All tones are compatible with lax vowels.
- Only tones 55, 33 and 21 are compatible with tense vowels.

Before we proceed to give a description of each layer of loanwords to Bai, we provide here some background on Chinese. Efforts to reconstruct Chinese have concentrated mainly on two periods: Middle Chinese (MC), a pronunciation system for Chinese characters embodied in the *Qie Yun*, a dictionary published in 601 C.E., which has good sound correspondences to modern dialects except Min; and Old Chinese (OC), the educated standard of China around 500 B.C.E. MC pronunciation is reconstructed by fleshing out the phonemic categories in the *Qie Yun* with sound values that can be regarded as ancestral to their reflexes in modern dialects. Although reference to modern dialects here is reminiscent of the comparative method, the MC categories are not derived through the comparative method. If the comparative method were applied to modern Chinese dialects, the result would presumably be a phonological system older than MC by several centuries. Such a system has not been reconstructed, however. The method for reconstructing Old Chinese is even more idiosyncratic: it takes advantage of the existence of two independent, yet convergent, bodies

² This paper uses Y. R. Chao's 5-level tone scale (from 5-high to 1-low). Two digits are used to mark the beginning and endpoint of a tonal contour; thus '21' means 'a tonal contour going from mid-low to low'.

of information: (a) the rimes in the Book of Odes, and (b) the phonetic element in the Chinese script. The reconstruction of OC morphology relies on internal reconstruction.

The phonological evolution between MC and the modern dialects, especially modern Mandarin, has been abundantly studied. Before proceeding further it is necessary to describe here the evolution of MC tones and manners of articulation into SW Mandarin, a variety of Mandarin spoken in the SW provinces of Guizhou, Sichuan and Yunnan:

<u>MC tones</u>	<u>Level 平</u>	<u>Rising 上</u>	<u>Departing 去</u>	<u>Entering 入</u>
<i>MC initials</i>				
<i>voiceless unasp. obstruents</i>	p-1	p-3	p-4	p-2
<i>voiceless asp. obstruents</i>	ph-1	ph-3	ph-4	ph-2
<i>voiced obstruents</i>	ph-2	p-4	p-4	p-2
<i>Sonorants</i>	m-2	m-3	m-4	m-2

Table 1: Reflexes of MC tones and manners of articulation in SW Mandarin, using labials to represent all places of articulation. 'p-1' in the first cell means that SW Mandarin normally has voiceless unaspirated obstruents in tone 1 corresponding to MC unaspirated obstruents under the MC Level tone.

We now give an outline description of each of the lexical layers in Bai (a detailed description of their phonological characteristics would require a monograph-size study). We begin with words obviously borrowed from southwestern Mandarin, which form two distinct layers: B1 and B2.

The local Mandarin layer B1

Judging from Middle Chinese (MC) tones and initial consonants, the disyllabic loans to Bai in this layer give the picture of a typical Mandarin dialect: the MC voiced obstruents have become voiceless aspirated under Level, but voiceless unaspirated under the other tones; the Level tone is split along the MC voicing distinction, the Departing tone is unsplit, and the Rising tone has lost its words with voiced obstruent initials to the Departing tone. What is noteworthy in a Yunnan context are the different reflexes for the MC lower Level and Entering tones, 21 and 35 respectively:

<u>MC tones</u>	<u>Level 平</u>	<u>Rising 上</u>	<u>Departing 去</u>	<u>Entering 入</u>
<i>MC initials</i>				
<i>voiceless unasp. obstruents</i>	33	21	<u>55</u>	35

<i>voiceless asp. Obstruents</i>	33	21	<u>55</u>	35
<i>voiced obstruents</i>	<u>21</u>	<u>55</u>	<u>55</u>	35
<i>Sonorants</i>	<u>21</u>	21	<u>55</u>	35

T0 = 33

Table 2: Reflexes of MC tones in B1 disyllabic loans to Bai (the lower tone series is in gray)

Jianchuan Mandarin is one of the few Mandarin dialects in Yunnan which maintain a distinction between the Entering and Lower Level tones. Here are the contours of Jianchuan Mandarin tones (based on Wu 1989:118): upper Level = 55, lower Level = 42, Rising = 31, Departing = 45, Entering = 21.³ Tones in Layer B1 and Jianchuan Mandarin are similar, especially for contour, except that the Entering tone is rising in B1. Examples (**Table 3**):

Gloss		tones in Jianchuan Mandarin ⁴	Bai
mother's brother	舅舅	D-D	tɕo ₅₅ tɕo ₅₅
beard	腮鬚	uL-IL	[lɕo ₅₅] sai ₃₃ xu ₂₁
crane	白鶴	E-E	pa ₃₅ xo ₃₅
chili	辣子	E-0	la ₃₅ tsi ₃₃
potato	洋芋	IL-D	ŋa ₂₁ jy ₅₅
woolen cloth	毛呢	IL-IL	mo ₂₁ ni ₂₁
head-cloth	包頭	uL-IL	po ₃₃ tho ₂₁
coral	珊瑚	uL-IL	sẽ ₃₃ xu ₂₁
spoon	調羹	IL-uL	thio ₂₁ kã ₃₃
means	辦法	D-R	pã ₅₅ fa ₃₅
trivet	三足	uL-E	sã ₃₃ tɕu ₃₅
two-stringed violin	二胡	D-IL	a ₅₅ xu ₂₁
dragon king	龍王	IL-IL	no ₂₁ uq̃ ₂₁
boundary	界限	D-D	ke ₅₅ q̃i ₅₅
dusk, twilight	黃昏	IL-uL	xuq̃ ₂₁ xuẽ ₃₃

³ The terms 'Level', 'Rising', 'Departing' and 'Entering' are traditional labels for Middle Chinese tone categories. They are without descriptive value in the modern dialects; the terms 'upper' and 'lower' denote their voiceless-initial and voiced-initial subsets, respectively.

⁴ Jianchuan Mandarin tone categories are here marked as: uL = upper Level, IL = lower Level, R = Rising, D = Departing, E = Entering, 0 = light tone (toneless syllable). A label such as 'upper Level' indicates a tone category originating in voiceless-initial subset of the Middle Chinese Level tone category.

future	將來	uL-IL	tɕã33 lɛ21
in the beginning	開始	uL-R	khe33 sa21
monday	星期一	uL-uL-E	ɕə̃33 tɕhi33 ji35
tuesday	星期二	uL-uL-D	ɕə̃33 tɕhi33 a55
honest	老實	R-E	lo21 sa35
arrogant, conceited	驕傲	uL-D	tɕo33 o55
polite	客氣	E-D	kha35 tɕi55
to keep secret	保密	R-E	po21 mi35
to sing a song	唱歌	D-uL	tshã55 ko33
to develop	發展	E-R	fɔ35 tsã21
to oppose	反對	R-D	fã21 tue55
to assemble/muster	集合	E-E	tɕi35 xu35
to pass, go by	經過	uL-D	tɕə̃33 kuɔ55
to queue	排隊	IL-D	pha21 tue55
to dance	跳舞	D-R	thio55 vv21
to prepare	準備	R-D	tsuẽ21 pi55

Table 3: Examples of B1 disyllabic words in Jianchuan Bai

This vocabulary is modern, but not very recent in character. One notes the American plants *chili* and *potato*, indicative of a Qing dynasty (1644-1911) date of borrowing. The "Cultural Revolution" vocabulary of Jianchuan Bai in Xu and Zhao (1984) is also clearly B1. Thus Jianchuan Mandarin is probably the source of B1 loans, and the period of borrowing extends at least from mid- or late Qing to the 1960s.

Not surprisingly, basic vocabulary items in this layer are very scarce: on a Swadesh-100 list, the only possible instance is 'claw' 爪子 *tsua21 tsɿ33*, which fits the B1 correspondences, although it could also belong to layer B2 (see below).

The regional Mandarin layer B2

The disyllabic loans in this layer point to a 4-tone Mandarin dialect with the same general Mandarin characteristics as B1, but here lower Level and Entering are merged. Absence of a distinction between upper and lower Level is probably not a feature of the source dialect, but the result of the impossibility for Bai speakers to reproduce the distinction using native Bai tones.

<u>MC tones</u>	<u>Level 平</u>	<u>Rising 上</u>	<u>Departing 去</u>	<u>Entering 入</u>
<i>MC initials</i>				
<i>voiceless unasp. obstruents</i>	55	21	<u>33</u>	55
<i>voiceless asp. obstruents</i>	55	21	<u>33</u>	55
<i>voiced obstruents</i>	55	<u>33</u>	<u>33</u>	55
<i>sonorants</i>	55	21	<u>33</u>	55

T0= 33 after 55, 21 after 33

Table 4: Reflexes of MC tones in B2 disyllabic loans to Bai (the lower tone series is in gray)

Yunnan Mandarin 4-tone systems are fairly stereotyped from the point of view of contours. The tones in the provincial capital Kunming 昆明 (Wu et al. 1989: 114) and in Heqing 鹤庆, a county prefecture adjoining Jianchuan in the East (Wu et al. 1989:118), are identical: Upper Level = 44, Lower Level, Entering = 31, Rising = 53, Departing = 213. That is the most standard type of tone system for Yunnan Mandarin. We will assume that a slightly different version of this system, in which the Lower Level, Entering category was mid-level 33 rather than mid-to-low falling 31, is the source of the B2 loans. Bai would then naturally have used its highest level tone, 55, to render the donor language's highest level tone, 44; it would have used its non-creaky falling tone, 21, to render the donor's only falling 53 tone; having no dipping tone of its own, it would have rendered the donor's 213 using its mid-level tone, 33. Finally, Bai would have been unable to distinguish between the donor's level tones, 44 and 33, treating them both as 55.

Examples:

Gloss		Yunnan Mandarin ⁵	Bai
Steam	蒸汽	uL-D	tsã55 tçhi33
Sulfur	硫磺	IL/E-uL	lio55 xuã55
business	生意	uL-D	sã55 ji33
Friend	朋友	IL/E-R	phã55 jo21
Buddhist nun	尼姑	IL/E-uL	ni55 ku55
Aunt	姨姨	IL/E-IL/E	ji55 ji55
Yak	牦牛	IL/E-IL/E	ma55 nio55
grape	葡萄	IL/E-IL/E	phu55 tho55

⁵ General Yunnan Mandarin tones are here marked as in **Table 3** above, with the difference that the merged lower Level + Entering category is marked as IL/E.

banana	芭蕉	uL-uL	pa55 tɕo55
tangerine	桔子	E-0	tɕu55 tsi33
cotton	棉花	IL/E-uL	mi55 xuɑ55
butter	酥油	uL-IL/E	su55 jo55
satin fabric	緞子	D-0	tuã33 tsi21
hat	帽子	D-0	mɔ33 tsi21
socks	襪子	E-0	va55 tsi33
boots	靴子	uL-0	ɕue55 tsi33
treasured	寶貝	R-0	po21 pe33
stool, bench	板凳	R-D	pa21 tã55
capital	本錢	R- IL/E	pã21 tshẽ55
interest	利息	D-E	li33 ɕi55
scissors	剪刀	R- uL	tɕi21 ta55
wheel	輪子	IL/E-0	nue55 tsi33
pack rack	架子	D-0	tɕa33 tsi21
story	故事	D-D	ku33 si33
joke	笑話	D-D	ɕo33 xuɑ33
fortune, luck	運氣	D-D	ɲue33 tɕhi33
temper, character	脾氣	IL/E-D	phi55 tɕi33
mark, sign	記號	D-D	tɕi33 xo33
color	顏色	IL/E- IL/E	ɲi55 sa55
zero	零	IL/E	ɲi55
cheap	便宜	IL/E- IL/E	phi55 ɲi33
pleasantly cool	清涼	uL-IL/E	tɕhã55 nia55
honest	規矩	uL-R	kue55 tɕy21
careful	細心	D-uL	ɕi33 ɕã55
happy and excited	喜歡	R-uL	ɕi21 xuã55
safe	平安	IL/E-uL	phiã55 ɲa55
affectionate	親熱	uL-E	tɕhã55 za55
clear up (liquid)	澄清	IL/E-uL	tɕ33 tɕhã55
transmit (to posterity)	傳代	IL/E-D	tshue55 te33
promise, consent	答應	E-uL	ta55 ɲa33
divide family	分家	uL-uL	fã55 tɕa55
separate	分開	uL-uL	fã55 khe55
complain to superior	告狀	D-D	ko33 tsuã33

assess, estimate	估計	uL-D	ku21 tɕi33
shy	含羞	IL/E-uL	xã55 su55
to regret	懊悔	D-R	o33 xue21
to doubt	疑心	IL/E-uL	ni55 ɕã55
drive car	開車	uL-uL	khe55 tshe55
consult, discuss, negotiate	商量	uL- IL/E	sã55 niã55
notify, inform	通知	uL-uL	thõ55 tsa55
want	想要	R-D	ɕã21 ɲo33
digest	消化	uL-D	ɕo55 xuɔ33
fight, vie for	爭搶	uL-R	tsẽ55 tɕhã21
turn a corner	轉彎	R-uL	tsuẽ21 ɲũe55
mule	驢子	IL/E-0	lo55 tsi33
donkey	驢子	IL/E-0	li55 tsi33
centipede	蜈蚣	IL/E-uL	ŋo55 kõ55

Table 5: Examples of B2 loans to Bai

B2 loans are about twice as numerous as B1 loans on our data. They are slightly more modern and urban in character: *butter*; *scissors*; *drive car*; *capital* (financial term); *interest*; *clothes*; there are fruit names (*banana*, *grape*, *tangerine*) and domesticated animal names (*donkey*; *mule*) but no plant names. We conjecture that the source of B2 loans is ‘standard’ Yunnan Mandarin, perhaps as spoken in Jianchuan county prefecture.

Basic vocabulary items in this layer are no more numerous than in B1: aside from ‘claw’, already mentioned, no Bai item in a Swadesh-100 list fits the B2 correspondences.

The early Chinese layer A.

We regard this layer as entirely borrowed from Chinese, like B1 and B2. This view will be justified in the rest of this paper. The disyllabic loans in this layer point to a non-Mandarin donor: the MC voiced stops are represented by unaspirated stops under each MC tone; the Level and Entering tones are only partially split; the Rising tone is unsplit; part of the Departing tone is represented by a separate tone; another part of it is identical with pre-split Entering.

<u>MC tones</u>	<u>Level 平</u>	<u>Rising 上</u>	<u>Departing 去</u>	<u>Entering 入</u>
<i>MC initials</i>				

<i>voiceless unasp. obstruents</i>	55	33	21 (some 33)	33
<i>voiceless asp. obstruents</i>	55	33	21 (some 33)	33
<i>voiced obstruents</i>	42 (some 55)	33	21 (some 33)	33, 21
<i>sonorants</i>	42 (some 55)	33	21 (some 33)	33, 21

T0= 55 (word-initially, ex.: 'tadpole'), word-finally 33

Table 6: Reflexes of MC tones in layer A disyllabic loans to Bai (the lower tone series is in gray).

In our original conference paper (see fn. 1) we gave full correspondences for initial consonants and rimes for this layer. However, as far as words of two syllables or more are concerned, in practice tone and initial consonant correspondences are enough for assignment to one or the other of our three layers.

In the table below, we give examples of layer A di- and polysyllables. Although layer assignment in monosyllables is more hazardous than in longer forms, we have added monosyllabic morphemes belonging to closed sets, such as the four seasons, the twelve-year cycle etc., when the entire set shows layer-A correspondences. In such cases the principle of extended coherence applies to the closed set paradigm instead of to a disyllabic morpheme.

Gloss		Middle Chinese ⁶	Bai
light	日照	IE-uD	ɲi33 tso33
moon	明月	IL-IE	mi55 ηuɔ33
weather	天日	uL-IE	xẽ55 ɲi33
thunder	天鳴	uL-IL	xẽ55 ma42
sea	大湖	ID-IL	tɔ21 ko42
dry fields	田地	IL-ID	xẽ55 tɕi21
paddy fields	水地	uR-ID	ɕy33 tɕi21
sand	沙子	uL-0	so55 tsi33
wave	波浪	uL-ID	po55 no42
household	人間	IL-uL	ɲi42 kã55
eyebrow	眼眉	IR-IL	ηue33 mi55
ear	耳頭	IR-IL	ɲi33 tiə42 [kuã55]
thigh	股頭	uR-IL	kuɔ33 tiə42 [kã33]

⁶ Historical tone categories are marked as uL 'upper Level', IL 'lower Level', uR 'upper Rising', IR 'lower Rising', uD 'upper Departing', ID 'lower Departing', uE 'upper Entering', IE 'lower Entering'.

calf of leg	細腳	uD-uE	se21 k _o 33
elbow	手肘子肘	uR-uR-uR-uR	si33 tse33 tsi33 tse33
finger	手指	uR-uR	si33 tsa33
thumb	手頭拇	uR-IL-IR	si33 tiə42 mo33
middle finger	中手頭指	uL-uR-IL-uR	tsō42 si33 tiə42 tsa33
brain	腦髓	IR-uR	no33 ɕy33
bone	骨頭	uE-IL	ku _o 33 tiə42
joint	手節腳節	uR-uE-uE-uE	si33 tse33 k _o 33 tse33
uvula	細舌	uD-IE	se21 tse21
common people	百姓	uE-uD	pa33 ɕã21
doctor	葯生	IE-uL	jo33 sã55
carpenter	木匠	IE-ID	ŋo33 tɕō21
blacksmith	鐵匠	uE-ID	the33 tɕō21
beggar	要餐飯	uD-uL-ID	ŋo33 tsã55 pē33 [xo33]
lame [person]	腳缺	uE-uE	k _o 33 khe33
son-in-law	女婿	IR-uD	ŋə33 so21
grandson	子孫	uR-uL	tsi33 suã55
buffalo	水牛	uR-IL	ɕy33 ŋə42
calf	細牛	uD-IL	se21 ŋə42
pony	馬駒子頭	IR-uL-uR-IL	ma33 tɕy33 tsi33 tiə42
lamb	細羊子	uD-IL-uR	se21 ŋo42 tsi33
dog	犬	uR	khuã33
deer	大鹿	ID-IE	to33 vu33
yellow weasel	鼠狼	uR-IL	su33 lo42
sparrow	雀子	uE-0	ts _o 33 tsi33
crow	黑烏	uE-uL	xə33 o55
tadpole	蝌蚪	uL-uR	ku55 tiə33
nit	白虱	IE-uE	pa21 ɕi33
silkworm	蠶子	IL-0	zã42 tsi33
poplar	水柳	uR-IR	ɕy33 ɣə33

cypress	松柏樹	uL-uE-ID	ɕõ55 pa33 tsi21
foodstuff, grain	五穀	IR-uE	ŋo33 ko33
awn of wheat	麥芒子	IE-IL-0	mə33 mo55 tsi33
millet	細白米子	uD-IE-IR-0	se21 pa21 me33 tsi33
soya bean	白豆	IE-ID	pa21 tiə21
powdered sugar	砂糖	uL-IL	so55 to42
soup	湯	uL	xã55
wheat bran	麥皮子	IE-IL-0	mə33 pe42 tsi33
pillow	枕頭	uR-IL	tsã33 tiə42
flight of steps	階臺	uL-IL	ka55 tiə42 [pã42]
mirror	鏡面	uD-ID	ka21 mi21
coal	火炭	uR-uD	xue33 thã21
dye	染料	IR-ID	zẽ33 lio33
frying wok	炒菜	uR-uD	tshu33 tshi21 [tshã55]
fire tongs	火鉗	uR-IL	xue33 tci42 [pa21]
section of bambo	竹筒	uE-IL	tso33 thõ55
saddle	馬鞍	IR-uL	ma33 õ55
girth	馬肚帶	IR-IR-uD	ma33 tu33 te33
manger	馬槽	IR-IL	ma33 tsu42
reins	牽馬	uL-IR	khẽ55 ma33 [sõ33]
carpenter's ink marker	墨斗	IE-uR	mə33 tiə33
sieve, sifter	籬頭	IL-IL	lo42 tiə42
gunpowder	火藥子	uR-IE-0	xue33 jo33 tsi33
trap	坑眼	uL-IR	khuõ21 ŋue33
character, word	書字	uL-ID	si55 tsi21
book	書冊	uL-uE	si55 tshua33
dragon king (god of rain)	龍王	IL-IL	no42 uã55
Buddha	佛	IE	ve21
physical strength	氣力	uD-IE	tchi33 ye21

birth day	生日	uL-IE	sã55 za21
age	日歲	IE-uD	ɲi33 suq33
danger	危險	uL-uR	ue55 ɕi21
use	用處	ID-uR	ɲõ21 tshu21
east	東	uL	tõ55
south	南	IL	na42
west	西	uL	sẽ55
north	北	uE	pə33
lower part of	下面	IR-ID	ɣa33 mi21 [no33]
under the sky	天下	uL-IR	xẽ55 ɣa33
tomorrow	明日	IL-IE	mẽ55 ɲi33
midnight	半夜	uD-ID	pã21 jo21
rat (year)	鼠	uR	su33
ox (year)	牛	IL	ŋə42
dragon (year)	龍	IL	no42
horse (year)	馬	IR	ma33
ram (year)	羊	IL	ɲo42
monkey (year)	猴孫	IL-uL	ŋo42 suã55
chicken (year)	雞	uL	ke55
dog (year)	犬	uR	kuã33
next year	後歲	IR-uD	ɣə33 suq33
spring	二三月	ID-uL-IE	za21 sã55 ɲuq33
summer	夏月	ID-IE	ɣo21 ɲuq33
autumn	秋月	uL-IE	tɕhə55 ɲuq33
winter	冬月	uL-IE	tõ55 ɲuq33
new year's day	新歲	uL-uD	ɕi55 suq33
1	一	uE	ji33
3	三	uL	sã55
4	四	uD	ɕi33
5	五	uR	ŋo33
6	六	IE	fɥ33
7	七	uE	tɕi33
8	八	uE	piq33
9	九	uR	tɕə33
10	十	IE	tɕa21

100	百	IE	pa ₃₃
1,000	千	uL	tɕhĩ ₅₅
10,000	万	ID	ŋua ₂₁
ordinal marker	第	ID	ti ₂₁
spacious	空寬	uL-uL	khõ ₅₅ khu ₃₃
narrow	窄狹	uE-IE	t _s a ₃₃ ka ₂₁
square	四面四角	uD-ID-uD-uE	ɕi ₃₃ mi ₂₁ ɕi ₃₃ ko ₃₃
expensive	價大	uD-ID	ka ₂₁ to ₂₁
young	日歲細	IE-uD-uD	ŋi ₃₃ su ₃₃ se ₂₁
clean	乾淨	uL-ID	kã ₅₅ tɕẽ ₂₁
near, close	隔近	uE-IR	ka ₃₃ tɕĩ ₃₃
worship	拜佛	uD-IE	pa ₂₁ ve ₂₁
punch hole	穿眼	uL-IR	tshue ₅₅ ŋue ₃₃
pierce through	戳破	IE-uD	tɕha ₃₃ pho ₂₁
to hammer in a nail	釘釘	uD-uL	tɕã ₂₁ tɕã ₅₅
to hide oneself	避藏起	ID-IL-uR	pi ₃₃ tsõ ₄₂ khẽ ₃₃
have a fever	發熱	uE-IE	fa ₃₃ ŋi ₃₃
worry, be anxious	惡心惡肝	uE-uL-uE-uL	o ₃₃ ɕĩ ₅₅ o ₃₃ kã ₅₅
set on fire	种火	uD-uR	tsõ ₂₁ xue ₃₃
to mix powder with water	和泥	IL-IL	yo ₂₁ ni ₄₂
avoid certain foods	忌嘴(1411)	ID-uR	tɕi ₃₃ tɕy ₃₃
immerse, submerge	浸入(1446)	uD-IE	tɕĩ ₂₁ ŋi ₃₃
chop down (tree)	割樹	uE-ID	tso ₃₃ tsi ₂₁
Understand	明白	IL-IE	ma ₄₂ pa ₂₁
Solidify	凝起	IL-uR	ŋẽ ₄₂ khẽ ₃₃
have shot (the target)	中得	uD-uE	tsõ ₂₁ tiẽ ₃₃
to grow	高大	uL-ID	ko ₅₅ to ₂₁
to rust	[]鐵鏽	uE-uD	[ta ₄₂] the ₃₃ sa ₃₃
to rise, go up	升起	uL-uR	sẽ ₅₅ khẽ ₃₃

put in order	收拾	uL-IE	si55 sa33
get dark	天暝	uL-IL	xē55 mia42
Hear	听得	uL-uE	tçhã55 tiə33
refuse by making excuses	推託	uL-uE	thue55 thuə33
be careful	細心	uD-uL	se21 çĩ55
welcome, greet	迎接	IL-uE	ŋɑ42 tçɑ33
increase, gain	加起	uL-uR	tçɑ55 khə33

Table 7: Examples of A-layer loans to Bai

Phonological characteristics of the A layer

Instability of correspondences over a long period of continuous borrowing

We define a layer of loanwords as *the set of all the loanwords borrowed in the course of a continuous contact period, however long, between two languages*. In the case of relatively short contact periods with intense borrowing, neither language normally has time to change significantly, resulting in compact layers with neatly storable rules of correspondence. Layers B1 and B2 are good examples. In the case of long contact periods, however, although one expects to see *some* continuity and stability in the sound correspondences, it is normal for both the languages involved in the contact relationship to undergo substantial phonological change during the course of the contact period. As a result the sound correspondences will change over time *within the layer*, defining a succession of sublayers. Typically, however, the boundaries between these sublayers cannot be drawn neatly, because the sound changes in both the donor and recipient languages will not be synchronized, and the more changes are taken into consideration when working out sublayers, the more complex and elusive the stratification becomes. Yet some aspects of the chronology of changes can often be recovered (see below).

Layer A is a good example of a layer of borrowings acquired over a long contact period, with complex sound correspondences evolving over time. There is an element of continuity in the correspondences (thus, for tones: upper level 55; Rising 33; upper Entering 33 throughout): but this layer has more unstable correspondences than the two Mandarin layers, with variation between tones, initial consonants and rhymes. This can be illustrated by the triplet for 二 'two': *ne33, ni21, za21*: these three forms are all part of layer A in our analysis (there is also a reading in layer B1: *ə55*, in 'Tuesday' and 'two-stringed violin'). From a phonological

standpoint, the sequence *ne33* > *ni21* > *za21* recapitulates the history of the word 'two' from Late Archaic to Late Mediaeval times: *ne33*, with its tense vowel and 33 tone, argues for a source form ending in a voiceless obstruent, perhaps final -s or -ts in the late OC pronunciation of 二 ^b*ni[j,t]-s*;⁷ *ni21* is close to the Middle Chinese pronunciation **nyijH* (Baxter 1992; 'H' indicates the Departing tone); and *za21* is close to the Late Middle Chinese pronunciation **ri* in Pulleyblank's reconstruction (Pulleyblank 1991), supposing the high front vowel had already been centralized under the influence of the retroflex initial in the donor Chinese dialect. The Bai forms for 'two' are further discussed below (Table 11 and text). Future research may make it possible to better sort out the different sublayers within layer A, but for the time being it will be sufficient to show that layer A is at least distinct from the modern layers B1 and B2.

One local Chinese donor or several?

At first sight the relative instability of correspondences might suggest that our A layer was borrowed not from one local variety of Chinese, but from successive forms of standard Chinese with which Bai was in contact at different periods. However, this would not account for the element of continuity in the representation of the Chinese tones throughout the duration of layer A, because the dialect base of successive Chinese standards during the period underwent several important shifts: the continuity can only be explained by supposing a local Chinese dialect whose tone system remained relatively stable during the period of contact, while it was itself becoming stratified through continuous contact with successive varieties of standard Chinese, as is the case with most, if not all, directly observable Chinese dialects.

Dating layer A

Chinese was introduced into Yunnan under the reign of Emperor Wu Di of the Han dynasty, in the late 2nd century B.C.E. This would presumably have been standard Western Han Chinese, a language of which little is known but which was perhaps based on, or at least influenced by, the speech of the western Han capital Chang'an (present-day Xi'an).

We can use linguistic data to gain an understanding of the upper date for the A layer by reference to events in Chinese phonological history: Old Chinese **l-* and **hl-* had changed to dental stops *d-* and *th-*, or to palatal continuants *j-* and *ç-* (depending on syllable type), by the

⁷ Old Chinese reconstructions are cited in the Sagart (1999) system. All reconstructed forms begin with superscript 'a' or 'b', to indicate type A or type B syllables.

end of the 1st century C.E. (Sagart 1999), while Old Chinese *r- did not change to l-, filling the gap left by the first change, until later (Ferlus 2005 argues for a 4th-century C.E. date). We find in Bai no examples of laterals corresponding to OC laterals. In type A syllables, OC words with lateral initials show dental stops reflexes: t- for OC *ʎ- as in 'peach' 桃 *ta42*, 'platform' 臺 *tiə42*, 'ground' 地 *tɕi21* (ti > tɕi through secondary palatalization) and in the ordinal marker 第 *ti21*; corresponding to OC *ʎhl- we find th- in 'iron' 鐵 *the33*, 'hear' 聽 *tchã55* (thia- > tcha- through secondary palatalization). Another set of words with OC lateral initials has x- in Bai: 'sky' 天 *xə55*, 'soup' 湯 *xã55*. The word 'field' 田 *xə55* shows x- corresponding to *ʎ-, seemingly suggesting that the donor Chinese dialect had *ʎhl- in this word (MC den implies OC ʎ-). We will return to the reflexes of OC hl- later. In type B the most common reflex of OC *bl- is j-: 'oil' 油 *jə42*, 'shake' 搖 *ju42*. One example shows ɲ-: 'use' 用 *ɲõ21*, reminiscent of the Yao form nlong (tone C2) 'to use' (Theraphan 1993). We have no clear examples of OC *hl- words in type-B words. This pattern of reflexes suggests that the Chinese dialect that Bai first came into contact with had already changed its laterals to other sounds. However it does not necessarily mean that contact began only after the changes *l- > d- and *hl- > th- in late Han (see Sagart 1999: 30-31): it may be that contact began in Western Han with a variety of Chinese which treated *ʎhl- as x-, and that the absence of examples of other kinds of laterals is accidental.

At the same time we find in Bai forms which represent OC *r- as ɣ-: 'strength' 力 *ɣə21*, 'come' 來 *ɣə21*, 牢 *ɣu21*, 樓 *ɣəu42*, 漏 *ɣə21*, 柳 *ɣə33*. Preceding i and u, OC *r- is represented in Bai by j- and v- respectively: 'beneficial' 利 *ji21*, 'chestnut' 栗 *ji21*; 'deer' 鹿 *vɹ21*; 'dredge for' 撈 *vu42*. OC *r- is represented as Bai l- in later layers ('old' 老 *lu33*, 'rite' 禮 *li33*), as it is in MC and most modern dialects. Thus the first Chinese dialect that Bai borrowed from still had a rhotic for OC *r-: the change to l- had not yet taken place. These two elements, the fate of OC *l- and *r- in early loans to Bai, indicate a lower date no later than the 1st century C.E., when OC laterals had already changed to their MC values, while OC *r- was still a rhotic. That Bai was already borrowing from Chinese by the end of the Han dynasty is confirmed by the fact that Bai maintains a distinction between OC *-u and *-aw in type A, as seen by Starostin (1995b):⁸ 槽 *tsu42*, 草 *tshu33*, 撈 *vu42*, 抱 *pu33*, 早 *tsu33* (< OC *-u) vs. 毛 *ma42*, 高 *ka55*, 桃 *ta42*, 刀 *ta55*, 盜 *ta21* (< OC *-aw). In Middle Chinese, this distinction is lost, but it was still observed in Eastern Han rhyming (Luo & Zhou 1958).

⁸ Starostin overinterpreted this retention as a sign of a close genetic relationship between Bai and Chinese.

An approximation for the time at which the A layer ended can be given based on the most innovative features in Chinese phonology to be found in layer A: change of labial stops to labiodental fricatives ('labiodentalization':) 'Buddha' 佛 *və21*, 'belly' 腹 *fɹ33*, 'fly' 飛 *fa55*; change of Shang-tone words with obstruent initials to Departing tone 21: 'house' 戶 *xo21*; 'danger' 危險 *ue55 əi21*. These changes are characteristic of Late Middle Chinese as opposed to Early Middle Chinese: they can hardly have entered Bai before mid-Tang (c. 750-800 C.E.). This situation testifies to the continuation of Bai-Chinese contacts during the Nanzhao Kingdom years (648-937). We have found no clear signs that linguistic interaction went on during the Dali Kingdom years (938-1253), however.

Other conservative features in the early part of the A layer

In the preceding section, we established an upper date for the A layer based on conservative features of the early part of that layer: having x- for OC *ʰl-; having ɣ-, or v-, or j- (depending on vocalic context) for OC *r-; and maintaining a distinction between OC -u and -aw in type-A syllables. Here we add a few more (a full discussion of phonological stratification with layer A would require a monograph-size study):

- absence of palatal semivowel medial in 'division-3' words. Exx.: 'drink' 飲 *ə33*, 'bridge' 橋 *ku42*. Contrast 'lamb' 羊 *no42*, 'immerse' 浸 *təi21*, 'water' 水 *əy33* later in the layer. Distinguishing the Div.-3 medial from a palatal main vowel is not straightforward, however.
- Lower Level and lower Entering treated as upper Level and upper Entering: 55 and 33. Exx.: 'wheat awn' 麥芒 *mə33 mo55*, 'moon' 明月 *mi55 ŋua33*. Contrast 'understand' 明白 *ma42 pa21* with distinct lower Level and lower Entering 42 and 21 later in the layer.
- retention of final -s in part of the Departing-tone category, leading to its treatment as Entering (33). Exx.: 'expensive' 價大 *ka21 to21*; contrast 'use' (n.) 用處 *no21 tshu21*, 'midnight' 半夜 *pə21 jo21* later in the layer.
- conservative vocalism:
 - OC *-ə still -ə: 起 *khə33*.
 - OC *-u still -u: 槽 *tsu42*, 草 *tshu33*, 撈 *vu42*, etc. (above)

We will speak loosely of an early sublayer with conservative features, a late sublayer with innovative features and a middle sublayer where the two types of features overlap. The boundaries between these three sublayers cannot be defined strictly, but we give below (Table 12) a chart of doublets which can illustrate the internal stratification of Chinese loans to Bai, including the internal stratification of the A layer. We think of these three sublayers

loosely as corresponding to the Han-Wei-Jin, Nanbeichao-early Mediaeval, and late Mediaeval periods respectively.

Characterizing the A-layer donor

We have argued that the Chinese donor language was a local (Yunnan) variety of Chinese which gradually became stratified through contact with successive varieties of standard Chinese. There are in layer A some interesting phonological features which may serve as clues in characterizing that language.

- *^ahl- > x- in 'soup', 'sky'

The existence in layer A of two reflexes: th- and x- for OC *hl- is interesting. The situation is similar to Middle Chinese, where one has e.g. 隋 'shred sacrificial meat' read as *^thwaX and *^xwjieH, in GSR 11, a clearly lateral series. In Sagart (1999), the normal MC reflex of OC *^ahl- was considered to be th-, while MC x- was treated as OC *^aq-hl-, with *q- an empty prefix: thus the two readings of 隋 would be reconstructed as *^ahloj? vs. *^bq-hloj(?)^{-s}. Sagart did not express much confidence in his reconstruction of a q- prefix, however:

'The evidence for this prefix is less abundant and varied than for the other prefixes; moreover, its functions have not been established. For this reason, I adopt it tentatively, as a measure making it possible to account for Middle Chinese reflexes while expressing *xie-sheng* and word-family connections.' (Sagart 1999:116).

Another explanation for the duality of reflexes is dialectal: OC *^ahl- would be reflected by th- in one dialect, and by x- in another. It is well-known that the word 'sky' 天 had two readings in Late Han/Nanbeichao times, one beginning in x- and another in th-. Their geographical distribution was described in Liu Xi's *Shi Ming* as eastern th- versus western x-, as recognized in Baxter (2005b). Given the geographical location of Bai, it makes sense that the local Han-time Chinese dialect would have pronounced Western features. That Bai does have x- in 'sky' and in two more words with OC laterals where Middle Chinese has th- ('field', 'soup') supports the idea that the evolution of OC *^ahl- to MC x- is a western dialectal feature.

- Reflection of OC -r as -n

Baxter (2005b) identified the reflection of OC -r as another pre-MC phonological feature which shows east-west variation: the evolution was to -j in Shandong and adjoining eastern dialects, while in the center and West the evolution was to -n. An example of evolution of -r to -j is in 'west' 西, MC *^sej < OC *^as-nər (Sagart 2004, there written as *^as-nəl). The Western

equivalent would be **asen*. Now the Bai word for 'west' is *sẽ55*, with nasal vowel, agreeing well with our hypothetical western form **asen*. This is another argument in favor of the idea that the local Chinese dialect had a western component.

We now discuss a couple of noteworthy forms:

- 'thigh' 股 *kua33* indicates OC **akwa?* rather than **aka?* (for independent evidence of this see Baxter 2005a; Feng 2005);
- 'middle' 中 *tsõ42* (in 'middle finger') with tone 42 (Lower Level) pointing to a voiced initial, while MC **trjuwng* and all modern dialects indicate a voiceless initial. Note that 'middle' in 'middle finger' is a stative verb: it is most likely that in the Chinese form underlying the Bai loanword, the voiceless initial, OC **btr-*, had been voiced by the intransitive N- prefix (Sagart 2003 for a recent account). Such a form, with Level tone and voiced initial, unknown from modern Chinese dialects and MC lexica, agrees exactly with the Chinese loan to proto-Hmong-Mien **ŋtɕuŋA* 'middle' (Wang & Mao 1995), where the nasal prefix is still visible.⁹

A special tone correspondence in layer A

Five tone-A morphemes with sonorant initials in layer A: 'person' 人 *ŋi21*, 'wolf' 狼 *ŋq21*, 'young person' 郎 *ŋq21*, 'shake' 搖 *ju21*, (animal) pen' 牢 *ɣu21*, have tone 21 instead of expected 42 or 55. The relevant items are listed in Table 8:

Adult	大人	<i>tõ21 ŋi21</i>
sick person	病人	<i>pã21 ŋi21</i>
Host	主人	<i>tsi33 ŋi21</i>
Others	人間	<i>ŋi21 kã55</i>
shake, quake	振搖	<i>tsẽ33 ju21</i>
sheep pen	羊牢	<i>ŋo42 ɣu21</i>
Wolf	狼	<i>ŋq21</i>
Girl	女郎子	<i>ŋə33 ŋq21 tsi33</i>

Table 8: Lower Level reflected as 21 in layer A

At least one of these morphemes: 'person' 人, has a segmentally identical variant with the expected 42 contour, in 'household' 人間 *ŋi42 kã55*. From the principle of extended coherence, it is clear that these items are part of layer A. Through the principle of coherence, the lower Level=42 feature can be ascribed to a sublayer of A which also reflects OC **r* as *ɣ-*

⁹ The Mienic reflexes shown tone C, probably from an *-s* suffixed Chinese form.

('pen'), and yet has developed a palatal medial in division-3 ('human being', 'shake'). The 21 contour could be a variant of the 42 contour of lower Level, but the conditions of variation are not understood.

Lexical characteristics of the A layer

Loanwords in this layer are most numerous, more than twice as numerous as in layer B2. It contains major lexical paradigms, such as:

- the numerals (though there are non-Chinese colloquial numerals for 'one' and 'two', see below);
- the four directions;
- the four seasons;
- the 12-year cycle.

In addition this layer contains:

- domestic animal names: the dog, cattle, horse;
- horsemanship vocabulary: saddle, girth, reins, manger;
- metal names: silver, copper, iron;
- cultivated plants: rice grain (but not 'paddy'), millet (again, the term indicates that this is a name for the grain, not the plant), wheat, peach, bean;
- wild animal names: the rat, wolf, jackal, fox, deer, yellow weasel;
- Buddhist terms: Buddha, stupa;
- names of artefacts: powdered sugar, gunpowder, vegetable oil, wheat flour, tea, medicine, pillow, bed, flight of steps, mirror, charcoal, chopsticks, plate or dish, cooking stove, lock, glue;
- names of tools: the axe, hammer, rope;
- a term for money.

With domesticated plants one notes the absence of American plants, consistent with an early date of borrowing. Note also that although there are names of domesticated plant and animals in this layer, the names of the rice plant, millet plant and pig, perhaps the most prominent targets of food production in the region are missing.

The term for 'gunpowder' is useful for datation purposes: gunpowder (a mixture of sulphur, saltpeter and charcoal) is believed to have been invented in mid-Tang. Overall the cultural contents of the layer is consistent with a Han to Late-Tang date (approx. 100 C.E. - 900 C.E.), in full agreement with the phonological characteristics of the layer.

The real surprise with this layer is the sheer amount of basic vocabulary it contains. On a Swadesh-100 list, we find as many as 47 items matching layer A: big 大 *tɔ21*, long 長 *tsõ42*, small 細 *se21*, daughter 女人 *ŋə33 ni21*, son 子人 *tsi33 ni21*, human (n.) 人間 *ni21 kã55*, dog 犬 *khuã33*, tree 樹 *tsi21*, seed 種子 *tsõ33 tsi33*, skin 皮 *pe42*, bone 骨頭 *kuq33 tiə42*, down, hair 毛 *ma42*, hair (of head) 頭毛 *tiə42 ma55*, eye 眼 *ŋue33*, hand 手 *si33*, belly 腹 *fu33*, heart 心 *ci55*, liver 肝 *kã55*, drink 飲 *ã33*, bite 咬 *ŋq33*, hear 聽得 *tchã55 tiə33*, swim 游水 *ŋã42 cy33*, come 來 *ɣə35*, sit 踞 *ku21*, speak 說 *suq33*, moon 明月 *mi55 ŋuq33*, star 星 *çã55*, water 水 *cy33*, sand 沙子 *so55 tsi55*, earth 土沙 *thu33 sa33*, cloud 雲 *vã42*, smoke 煙子 *ni55 tsi33*, fire 火 *xue33*, path 途 *thu33* (actually 'road'), red 赤 *tsha33*, green 綠 *lu33*, yellow 黃 *ŋo42*, white 白 *pa21*, black 黑 *xə33*, midnight 夜 *jo21*, full 滿 *ma33*, new 新 *ci55*, round 圓 *ŋue42*, dry 乾 *kã55*, fly (v.) 飛 *fa55*, flesh 肌 *ka42*, die 死 *ci33*.

However, two facts must be borne in mind:

- Only a small portion of these items (words like 'skin', 'hair' (of head), 'come', 'sit', 'fire') clearly belong to the *early* part of layer A; many, like 'big', 'long', 'small', 'woman', 'man', 'human being', 'tree', 'skin', 'feather', 'hand', 'belly' etc., have phonological characteristics indicative of the later part of the layer: even if Bai was most closely related to Chinese, these words could not be shared by Bai and Chinese as a result of common inheritance.
- There is also a sizeable number of non-Chinese, Sino-Tibetan-related basic vocabulary items in Bai, to be examined in the next section.

The Tibeto-Burman layer

The Bai lexicon contains a non-Chinese, Sino-Tibetan (henceforth 'ST') layer which we call Tibeto-Burman (henceforth 'TB') here for ease of reference, although the existence of a Tibeto-Burman branch of ST needs to be supported by a body of shared innovations. Given the lack of wholly explicit systems of reconstruction for either Tibeto-Burman or Sino-Tibetan (Sagart, 2006), it has not been possible to constrain our study of the TB lexicon in Bai using sound correspondences between Bai and a reconstructed TB or ST pronunciation. In Lee and Sagart (1998), we compared Jianchuan Bai words with Proto-Loloish reconstructions by D. Bradley (1979). We tentatively ascribed Bai words to the TB layer when (a) they could not be related to a Chinese etymon by means of the segmental and tonal correspondences extracted from our study of layers B1, B2 and A, and (b) when they showed some sound correspondences to Bradley's Proto-Loloish. We presented 39 comparisons between Bai and Proto-Loloish. We reproduce the list below, reduced to 25 comparisons (we have removed

comparisons which we now consider erroneous, or which can with equal plausibility be regarded as Chinese loanwords):

gloss	TBL	Proto-Loloish (Bradley)	Bai
rain	10	*r-ywa ¹	va33, za33
pig feed	459	*dza ¹	tsa33
l	928	*C-nga ¹	ŋo21
thou	931	*nang ¹	no21
water	158	*re ¹	ji21
(<tears)			
this	942	*no ¹ 'that' (near)	no21
nose	80	*s-na ¹	ne42
tiger	304	*k-la ²	lo42
sky (<dawn)	1671	*mo ²	ma42
speech,	651	*dang ²	tō42
words			
head	80	*u ²	vu42
(<nose)			
earring	484	*k-wang ²	kō42
pine torch	388	*C-mi ² ('fire')	me55
hard,	1095	*s-ra ²	so55
laborious			
urine	152	*ʃi ²	sa55
earth	40	*tʃa ²	sa33
blood	129	*swe ²	sug33
mother	219	*ma ³	mo33
to dance	1677	*ga ³	ko33
eye (<tears)	79/158	*C-myak ^H 'eye'	mi21
expensive	1054	*kak ^H	ka21
untie, undo	1600	*k-luat ^H 'free'	luq33
paddy rice	403	?-gok ^H	ko42
cold		?-grak ^H	ka42
enough	1349	*lok ^L	lu55

('<_' means 'morpheme extracted from the word for _')

Table 9: words shared by Bai and Proto-Loloish (Bradley 1979)

One notes terms for 'pig feed' and 'paddy rice', showing this layer originates in a more rural population than layers A, B1 or B2.

On the basis of the comparisons such as those in Table 9 and of sound correspondences we detected in them, we argued that Bai was probably Loloish. In a rejoinder, Matisoff (2001) offered a number of additional comparisons, arguing that when Jianchuan tones are replaced in a general Bai context, including other Bai dialects, the correspondences we detected between Bai and Loloish become less clear.¹⁰ We will reserve discussion of Bai vs. Loloish for another occasion.

We now identify Bai basic words on a Swadesh-100 list which we regard as being of probable TB origin:

1. **I (1sg):** *ŋo21*. This could represent the modern Mandarin 1sg 我 'I' in layer B1, with trivial correspondences, but if so, that would be the only basic word borrowed in that layer. The tone would have to be 33 if it belonged to layer A. This word is more plausibly compared to TB *nga*, PL *C-nga*₁ 'I'. If so, it is essentially regular.
2. **thou:** *no21*. Irregular vowel and/or tone as a reflex of any of the n- initial Chinese 2nd person pronouns: 汝, 乃, 爾. Straightforward comparison to PL **nan*₁, compare PL **dang*₁ 'speech, words', Bai *tõ42* 'id.'. Nasal vowels are regularly denasalized in Bai after nasal initials.
3. **we:** *ŋa55*. Could be a reflex of 我, a plural pronoun in OC, but vowel correspondence is unparalleled and Bai tone /55/ is irregular corresponding to Chinese Rising tone. Enters into a PL tone-2 correspondence with Wuding *ŋu11* and Sani *ŋe33* 'id.', therefore the word is clearly ST but probably non Chinese.
4. **this.** *no21*. Not a Chinese form. Compare PL **no*₁ 'that (near)'.
5. **not.** *a33 ~ ja33*. The two forms are variants. Neither is Chinese. With *a33*, compare the general Jingpo negation *a31*, and phonetically similar forms of same meaning in Yi (Zhao Yansun 1982: 165).
6. **one.** *a21*. A plausible cognate set for this item was proposed by Zhao (1982): Taoping Qiang *a31*, Aka (= Hruso) *a*, both 'one'; Xixia *a* 'one of a pair'.
7. **two.** *kõ33*. The connection to Jingpo *la55 khor51* 'two' was first pointed out in Zhao (1982). Add Sulung (Tayeng 1990; spoken in NW Arunachal Pradesh, a "central TB"

¹⁰ In his rejoinder, Matisoff refers to Lee & Sagart (1998) as "Sagart (1998)".

language according to Bradley's classification) *akuŋ*, 'second'¹¹, Rengma (eastern Naga) *koŋhu* '2' and Phun (Burmic) *naikoŋ*. Wang (2006:166) claims this Bai word is cognate to 兩 'two', OC *braŋʔ. However he cannot give any evidence that Bai ever had r- and that Chinese ever had k- in this word.

8. **blood.** *suŋ33*. The final and tone correspondences to 血 'blood', MC *xwet* are regular in the old layer but the initial correspondence is not. Better compared with PL *swe² 'blood'.
9. **breast(s).** *pa21 tci33*. Compare Wuding Yi *a55 pa2id*.'
10. **walk.** *pe33*. A Tai origin has been suggested but this is better compared to PL *p-re² 'run'.
11. **rain (n.).** *za33 ~ va33*. *za33* should be compared with PL *r-ywa¹, and *va33* with an unprefix variant of the same. A comparison with 雨 MC *hjo* is feasible for the initial and tone, but Bai /a/ corresponding to this Chinese rhyme (and generally to the OC 魚 rhyme category) is unparalleled.
12. **mountain.** *su21*. Not a Chinese word (山 MC *srean* cannot correspond). The word is better compared with Hani *fv31*, Naxi *dzy21* 'id.'

Although this list is quantitatively limited, it is more basic in character than the list of 47 Chinese-Bai matches on the Swadesh-100 list presented above: it contains in particular three personal pronouns, a demonstrative pronoun, and the lower numerals 'one' and 'two'.

Determining the genetic layer in Bai

Archaeologists derive relative chronologies of cultural layers based on stratigraphy: older cultural layers are lower in the ground than more recent ones. Are there sites in languages where successive language strata are stratified in an obvious fashion? we believe the numeral system of a language is one. It is very generally the case that when a language borrows some numerals under ten from another language, the borrowed numerals will form a continuous set beginning with 9 and ending with a lower numeral (such as 9-8-7-6 or 9-8-7-6-5-4-3). If any indigenous numerals survive, they will form another continuous set below the lowest borrowed numeral (such as 5-4-3-2-1 or 2-1).

In Bai, all numerals above 'two' are Chinese. For 'one' and 'two', however, we find two competing sets: a set of forms of Chinese origin: *ji33*, *ji35* 'one' and *ni33* 'two'; and a set of forms of TB origin: *a21* 'one' and *kõ33* 'two' (Table 10).

1.	<i>ji33</i> , <i>ji35</i>	<i>a21</i>
2.	<i>ni33</i> , <i>ne33</i>	<i>kõ33</i>

¹¹ We thank F. Jacquesson for bringing the Sulung datum to our attention.

3.	sã55
4.	ɕi33
5.	ŋo33
6.	fɯ33
7.	tɕi33
8.	piɣ33
9.	tɕə33
10.	tʂa21

Table 10. Bai numerals of Chinese origin (shaded cells) and of TB origin (clear cells)

The numeral *ji35* is a B1 loan used exclusively in spelling out numbers in modern contexts such as year names, telephone numbers, codes etc. (Xu & Zhao 1984: 24sq). In counting things, the Chinese numerals *ji33* 'one', *ne33* and *ni21* 'two' are limited to numbers above 10: eleven, twelve, twenty, twenty-one, twenty-two, thirty-two, two hundred, etc.; only the TB numerals *a21* and *kõ33* can be used as 'one' and 'two' in counting things. Here are examples drawn from Xu and Zhao (1984:24):¹²

ji21-kã22 a31 ji21
 person one CLASSIFIER
 “one person”

pẽ33 kõ33 jo21
 board two CLASSIFIER
 “two boards”

The Chinese-related numerals cannot be used in this most basic and colloquial function. This shows that the Chinese-related numerals, which belong to layer A, are borrowed, and that the TB numerals are inherited. In other words, layer A is borrowed and the TB layer is genetic. This fits well with our observation that the TB layer, although smaller in size than the any of the Chinese layers, is more basic in character.

Summary and conclusion

We have identified four chronological strata in the Bai lexicon. The genetic layer is Sino-Tibetan, and clearly non-Chinese. It contains the personal pronouns, the numerals for 'one' and 'two', and other items making up at least 12% of a Swadesh 100-word list. The cultural vocabulary ascribable to this layer includes words relating to rice cultivation and the raising of

¹² Tone values in Xu and Zhao are slightly different from those used in TBL.

pigs. The next layer (layer A) was borrowed from Chinese in a long and complex episode of intimate contact lasting from Han to late Tang. 47% of a Swadesh 100-word list, as well as enormous amounts of cultural vocabulary, mostly urban in character, were borrowed from Chinese in the course of that episode. Linguistic interaction between Bai and Chinese appears to have petered out in the first half of the second millennium C.E., to resume in late Ming or Qing times with the introduction of Mandarin to the Chinese southwest. Two distinct layers accommodate recent loans from two varieties of SW Mandarin: Jianchuan Mandarin (B1) and 'regional' SW Mandarin (B2). Almost no basic vocabulary is to be found in these two layers.

Let us illustrate the stratification of the Bai vocabulary using the word 'two' (Table 11) and other examples (Table 12). We believe the three forms of 二 'two' in layer A were borrowed from Chinese at different periods, as part of different lexical items: *ne33* early, as an independent word used in literary contexts and in counting objects above ten; *ni21* later on, exclusively in 'twenty' *ni21 ei21*, which must have displaced an earlier *ne33 tsa21*; *za21* in the late Middle Chinese period, in expressions like 二三月 *za21 sa55 ŋua33* 'spring'; and finally *a55* from Jianchuan Mandarin, in words like 'Monday' and 'two-stringed violin'. Note that 二 was never borrowed as an independent word, used for counting objects in twos. In that most basic function, Bai uses *kō33*, a non-Chinese form with cognates in Jingpo and other ST languages.

layer	Bai	source	use in Bai
TB	kō33	(genetic)	'two'
A _{early}	ne33	late OC ni[jt]s	cardinal in 'twelve', '(twenty)-two', 'one hundred and two', etc.; ordinal in 'second'.
A _{middle}	ni21	early MC nyij _{tone C}	in 'twenty'
A _{late}	za21	late MC ɲzi _{tone C}	'second month'
B2	None	SW Mandarin	
B1	a55	Jianchuan Mandarin	'Monday', 'huqin' (two-stringed-violin)

Table 11: Bai lexical stratification as illustrated by the words for 'two'. Data from Xu & Zhao (1984:24ff.)

layer	二	日	大	細	話	毛	盤	龍	分	十
A _{early}	ne33	ɲi33							pā55	
A _{middle}	ni21		to21	se21	yo21	ma42	pā42	no42		tsa21

A _{late}	za21	za21								ci21
B2			tq33	ci33	xuq33	ma55			fā5	
B1	a55						phā21	no21		

Table 12: More examples of Bai lexical stratification

There, precisely, lies the interest of Bai, a language which has borrowed almost half the words on a Swadesh-100 list from Chinese, while its genetic layer contains less than fifteen matches on the same list. To that extent, Bai is counterevidence to Starostin's claim (1995a: 395) that there are limits to lexical borrowing, specifically that a language cannot borrow more than 15 % of a Swadesh 100-word list. Starostin argued that once a language has reached that stage, its speakers will shift to the dominant language. Bai shows that this is not the case. The genetic layer in a language cannot be determined mechanistically by looking at the number of matches on a basic vocabulary list.

Bai may be considered one of the world's borrowing champions. How it could borrow so many items from Chinese in the approximately 1000-year period between early Han and Late Tang is not known. Presumably there was intimate contact, widespread bilingualism in Bai cities, probably also high levels of literacy in Chinese, combined with factors favoring the maintenance of Bai in the face of cultural pressure. *A contrario*, that almost no basic vocabulary was borrowed from Mandarin in the course of 700 years since the Yuan dynasty suggests that present-day generalized Bai-Mandarin bilingualism is recent.

REFERENCES

- Baxter, W. H. III. 2005a. “上古音脂、微两部之间的界限 [the boundary between the Zhi and Wei rhyme categories in Old Chinese]”. Power-Point presentation for the *First International Symposium on Old Chinese*, Shanghai, December 14-18, 2005.
- Baxter, W. H. III. 2005b. “Early Chinese dialects”. Power-Point presentation for the *First International Symposium on Old Chinese*, Shanghai, December 14-18, 2005.
- Benedict, P.K. 1982. “Chinese and Proto-Tai, part II”. Paper presented at the *15th international conference on Sino-Tibetan Languages and Linguistics*, Beijing 1982.
- Bradley, D. 1979. *Proto-Loloish*, Scandinavian Institute Monograph Series N39. London and Malmo: Curzon.
- Dell, F. 1981 *La langue Bai, phonologie et lexique*. Paris: Editions de l'Ecole des Hautes Etudes en Sciences Sociales.
- Feng, Zheng. 2005. “论上古音鱼 (铎) 部中模韵系的开合口问题 [On labiality in the Mo rime of the Yu-duo rhyme category]”. Paper presented at the *First International Symposium on Old Chinese*, Shanghai, December 14-18, 2005.

- Ferlus, M. 2005. "L'intérêt linguistique des transcriptions chinoises concernant le Cambodge ancien (Fou-nan et Tchen-la)", *19èmes journées de linguistique Asie Orientale*, Paris, 30 Juin-1er Juillet 2005.
- Fu, Jingqi and Xu Lin. 2001. "On the origin of seventeen Bai words found in Manshu". Paper presented at the *34th International Conference on Sino-Tibetan languages and linguistics*, Kunming, China.
- Huang, Bufan et al. 1992. *Zang-Mian Yuzu Yuyan Cihui [A Tibeto-Burman Lexicon]*. Beijing: Zhongyang Minzu Xueyuan Chubanshe.
- Lee, Yeon-Ju and Sagart, Laurent. 1998. "The strata of Bai". Paper presented at the *31st International Conference on Sino-Tibetan Languages and Linguistics*. University of Lund, September 30 - October 4, 1998.
- Li Fang-kuei. 1937. Languages and dialects of China. *The Chinese Yearbook, 1937*. Shanghai: Commercial Press.
- Li, Shaoni and J. A. Edmondson. 1990. Yunnan Jianchan Baiyu yinzhi he yindiao leixing [Phonetic and Phonological type of Jianchuan Bai of Yunnan]. *Zhengyang minzu xueyuan Xuebao* 5.
- Luo Changpei and Zhou Zumo. 1958. *Han-Wei-Jin Nanbeichao Yunbu yanbian yanjiu. [researches on the evolution of rimes in the Han-Wei-Jin and Nanbeichao periods]*. Beijing: Kexue Chubanshe.
- Matisoff, J.A. 2001. "On the genetic position of Bai within Tibeto- Burman". Paper presented at the *34th International Conference on Sino-Tibetan languages and linguistics*, Kunming, China.
- Norman, J. 1979. "Chronological strata in the Min dialects". *Fangyan* 4: 268-274.
- Pulleyblank, E. 1991. *Lexicon of reconstructed pronunciation in Early Middle Chinese, Late Middle Chinese and Early Mandarin*. Vancouver: University of British Columbia Press.
- Sagart, L. 1999. *The Roots of Old Chinese*. Amsterdam: John Benjamins.
- Sagart, L. 2004. "The Chinese names of the four directions". *Journal of the American Oriental Society*, 124, 1: 69-76.
- Sagart, L. 2006. "Review of J. A. Matisoff Handbook of Proto-Tibeto-Burman". *Diachronica* 2006.
- Sagart, Laurent & Xu Shixuan. 2001. "History through loanwords: The loan correspondences between Hani and Chinese", *Cahiers de Linguistique-Asie Orientale* 30(1):3-54
- Starostin, S.A. 1995a. "Comments from Starostin. In: William S.-Y. Wang (ed.) The Ancestry of the Chinese Language". *Journal of Chinese Linguistics monograph series* no. 8, 393-404.
- Starostin, S.A. 1995b. "The Historical Position of Bai". *Moskovskij Lingvisticheskij Zhurnal* 1, 174-190.
- Sun, Hongkai et al. 1991. *Zangmianyu Yuyin he Cihui [Tibeto-Burman phonology and lexicon]*. Beijing : Zhongguo Shehui Kexue Chubanshe.

- Tayang, Aduk. 1990. *Sulung language guide*. Itanagar: Directorate of Research, Govt. of Arunachal Pradesh.
- Theraphan, L.-Thongkum. 1993. "A view on Proto-Mjuenic (Yao)". *Mon-Khmer Studies* 22: 163-230.
- Wang, Feng. 2006. *Comparison of languages in contact: the distillation method and the case of Bai*. Institute of linguistics, Academia Sinica.
- Wang, Fushi and Mao Zongwu. 1995. *Miao-yao yu guyin gouni [a reconstruction of the sound system of Proto-Miao-Yao]*. Beijing: Zhongguo Shehui Kexue.
- Wiersma, Grace. 1990. *Investigation of the Bai(Minjia) language along historical lines*. UC Berkeley: PhD thesis.
- Wiersma, G. 2003. "Yunnan Bai". *The Sino-Tibetan languages*, ed. by Graham Thurgood & Randy J. LaPolla: 651-673. London: Routledge.
- Wu, Anqi. 2000. "Zangmianyu de fenlei he Baiyu de guishu. [the classification of Tibeto-Burman and the position of Bai]". *Minzuyuwen*, 1:1-12. Beijing.
- Wu, Jicai et al. 1989. *Yunnan Shengzhi. Hanyu Fangyan zhi [A Yunnan province gazetteer]. Section on Chinese dialects*. Kunming: Yunnan Renmin Chubanshe.
- Xu, Lin & Zhao Yansun. 1984. *Baiyu jianzhi [researches on Bai]*. Beijing: Minzu Chubanshe
- Zhang, Zengqi. 1990. "Zhongguo xinan minzu kaogu [Archeology on nationalities in South-western China]", *Yunnansheng bowuguan yanjiu congshu*. Kunming : Yunnan Renmin Chubanshe .
- Zhao, Yansun. 1982. "Baiyu de xishu wenti [The problem of the genetic relationship of Bai language]", *Minzu yuwen yanjiuji [collection of studies on Minority linguistics]*. 150-188. Xi'ning: Qinghai Minzu Chubanshe.
- Zhao, Yansun & Xu Lin. 1996. *Bai-Han cidian [Bai-Chinese dictionary]*. Chengdu: Sichuan Minzu Chubanshe .
- Zhengzhang, Shangfang. 1999. "Baiyu shi Han-Bai yuzu de yizhi duli de yuyan [The Bai language is an independent branch of the Sino-Bai Family]", *Zhongguo yuzhan: Qingzhu Wang Shiyuan jiaoshou liushiwu sui huadan*, 19-73. Xianggang: Chengshi Daxue Chubanshe.