SOAS
University of London

ARTICLE

Bai and Old Western Chinese

Lee Man Hei

The Chinese University of Hong Kong, Hong Kong
Email: 1155034402@link.cuhk.edu.hk

Abstract

Several scholars noted that the pronunciations of 天 "sky" *tiān* and 風 "wind" *fēng* in Bai appear to be akin to the western variants of the words attested in the paronomastic gloss dictionary *Shìmíng* 釋名. I will demonstrate in the current study that there are additional commonalities shared by both Bai and the ancient western dialect, termed Old Western Chinese (OWC) in this study. In both languages, one can identify words with *zy-* in Middle Chinese (MC) that are pronounced *j-*. Bai and Old Western Chinese use the same word (椹 *shèn*) for "fungus". Furthermore, Old Chinese (OC) cluster *-p/t-s yields -t in both languages in lieu of yielding -j as observed in Middle Chinese. Last but not least, it appears that in both languages, words with *lˤ- (whence MC *d-*) and -ʔ (whence MC rising tone) are distinct from other words with *d-* in Middle Chinese. Hence, this paper puts the claim that Bai is akin to Old Western Chinese on a stronger footing. As a side note, judging from the fact that 四 "four" *sì* contains -t in Old Western Chinese and early Bai, its Old Chinese form most likely ends in *-[t]-s.

Keywords: Bai; Old Western Chinese; Chinese; Old Chinese; Middle Chinese; Historical linguistics

1. Introduction

The Bai language is mainly spoken in northern Yunnan and is the native language of more than a million people. The classification of Bai is disputed; it has been claimed to be a Sinitic language (Greenberg 1953; Benedict 1982; Starostin 1995; Zhengzhang 1999), a Tibeto-Burman language (Matisoff 2001; Lee and Sagart 2008), or a sister group of Chinese (Wang 2013).[1] As with many languages in southwestern China, plenty of (Southwestern) Mandarin loanwords, which are typically non-basic words, have been introduced to the language in recent centuries. Nevertheless, unlike Loloish languages and Naic languages, there is also a myriad of words of Sinitic origin that are apparently

---

[1] Zhengzhang (1999) asserted that Sinitic and Bai constitute a branch within Sino-Tibetan called Sino-Bai. Later, he (Zhengzhang 2010) added Caijia to the Bai branch and postulated that these two languages originate from the Chu language 楚語 spoken in the late Warring States period. The reason Zhengzhang did not include Bai and Caijia in Sinitic is because they are fairly divergent from "Sinitic". Nonetheless, he did not explicitly state how he defined Sinitic. Most scholars in the field of historical Chinese phonology regard Old Chinese as the parent node of Sinitic and, based on Zhengzhang's view that the Bai branch derives from the Chu language, his proposed Sino-Bai group is in effect Sinitic, as the Chu language (in the late Warring States period) would normally be regarded as a variety or daughter language of Old Chinese in lieu of its sister language. Regarding Wang's (2013) work, the parent node of "Chinese" is presumably Old Chinese.

not recent loanwords. This layer of words is remarkably extensive, many of which are basic words. Starostin (1995) identified dozens of early Sinitic words, which include the majority of concepts in the Swadesh 100-word list for Bai. Zhengzhang (1999) identified a substantially larger number of early Sinitic words, as he examined a larger body of Bai words. Nevertheless, his analysis also appears to contain more errors; for instance, Mandarin loanwords are sometimes mistakenly treated as early Sinitic words, e.g. he identified $ɕã^{44}$ (as in $ɕã^{44}$ $pʰĩ^{44}$ "photo") as 相 MC $sjang^H$ > xiàng and placed it in the early layer, but this morpheme is apparently borrowed from Mandarin; Middle Chinese rhyme -jang corresponds to -õ in the early layer. Additional work is then needed to identify genuine early Sinitic words from his work. In this section and Sections 3.2–3.4, the early Sinitic words are selected from Starostin's and Zhengzhang's works: the correspondences between Bai and Middle Chinese identified by Starostin (1995) are used to isolate genuine early Sinitic words from Zhengzhang (1999). The current author holds the opinion that Bai is a Sinitic language and will delve into this issue in another study. For this reason, and as the current study is virtually always about the early Sinitic layer of Bai in place of recent Mandarin loanwords, the layer will generally be referred to as Bai hereafter for the sake of convenience.

Bai is notable for preserving some archaic words and conservative phonological features.[2] For instance, "red" is represented by 赤 OC *t-qʰrAk > chì – Bai $tsʰɛ^{44}$, "low, short" represented by 庳 OC *N-peʔ > bì – Bai pi^{33}, "sleep" represented by 寢 OC *tsʰimʔ > qǐn – Bai $tsʰɛ̃^{33}$, "firewood" represented by 薪 OC *sin > xīn – Bai $ɕĩ^{55}$, etc. These words are no longer the common terms for the respective meanings in Mandarin, Yue, Wu, Hakka, etc. In terms of phonology, some aspects of Bai's early Sinitic layer cannot be accounted for in terms of Middle Chinese. For instance, the distinction between OC *-u and *-aw/ew is retained in words with -aw or -aew in Middle Chinese (Starostin 1995: 182–3), as shown in examples (1) and (2). There are also innovative features that are probably not derived from Middle Chinese. For instance, OC *r- becomes ɣ- in certain Bai words (see example (3)) in lieu of l-, the reflex seen in Middle Chinese and mainstream modern varieties.

(1) 草 "grass" OC *tsʰˤuʔ > MC $tsʰaw^X$ > cǎo – Bai $tsʰu^{33}$
   桃 "peach" OC *C.lˤaw > MC daw > táo – Bai ta^{21}
(2) 飽 "satiated" OC *pˤruʔ > MC $paew^X$ > bǎo – Bai pu^{33}
   豹 "leopard" OC *pˤrewk-s > MC $paew^H$ > bào – Bai $pã^{42}$
(3) 來 "come" OC *mə.rˤək > MC loj > lái – Bai $ɣu^{35}$
   力 "strength" OC *k.rək > MC lik > lì – Bai $ɣu^{42}$

Scholars have touched upon to which historical variety Bai is related. Lee and Sagart (2008) linked the Bai word $xɛ^{55}$ "sky" to the western variant of 天 "sky" attested in some early works, including Shìmíng. In a similar vein, Gong (2015) associated the Bai word pi^{55} "wind" to the western variant of 風 "wind" attested in Shìmíng. I will demonstrate in this study that there are additional commonalities shared by Bai and the ancient western dialect, or Old Western Chinese. Thus, this paper consolidates the claim that Bai is akin to Old Western Chinese.

---

[2] The Old Chinese forms in this study are cited from Baxter and Sagart (2014) and the Middle Chinese forms are represented by Baxter's Middle Chinese transcription. Brackets in OC forms are omitted in most cases; they are retained when they are pertinent to the point in question. If not explicitly stated otherwise, the Bai forms are those of Jiànchuān 劍川 dialect cited from Zhao and Xu (1996).

2. Old Western Chinese

A number of features illustrated in the sections below are attested in both Sui/Tang and Song works. For instance, the rhyme dictionary *Qièyùn* 切韻, published during the Sui dynasty, states that (some) departing-tone words have a stop coda in Guānzhōng 關中 and Gānsù. In the Song rhyme dictionary *Jíyùn* 集韻, departing-tone words ending with -*t* could be found in Guānzhōng. Another example would be the Guānzhōng variant of 稻 "paddy" OC *lˤuʔ > MC *daw^X*. It is stated in both *Guóshǐ Bǔ* 國史補, completed in the Tang dynasty, and *Jíyùn* that 稻 was pronounced *thaw^X*, with initial *th*-.

Based on the continuity of the above features in Guānzhōng from the Sui/Tang dynasties to the Song dynasty, it is reasonable to posit that there is also continuity between the Guānzhōng varieties spoken in Sui/Tang times and that spoken in Song times. The exact geographical extent of the lect in the respective periods is difficult to determine, but it always includes the Guānzhōng basin, which is situated in the western part of China proper. As such, this historical variety, spoken between the Sui/Tang and Song dynasties, is termed Old Western Chinese in the present study.

There is a western feature that was attested not only in the Tang and Song dynasties but in the earlier Eastern Han dynasty as well. According to *Shìmíng*, there were dialectal variations in terms of how 天 OC *lˤin > MC *then* > *tiān* was pronounced. The word is pronounced as *tʰˤen in Qīngzhōu 青州 and Xúzhōu 徐州 (large parts of Shāndōng and Jiāngsū) but as *xˤen in Yùzhōu 豫州, Sīzhōu 司州, Yǎnzhōu 兗州, and Jìzhōu 冀州 (Central Plain and Guānzhōng basin).[3] The key difference between the two variants lies in the onset, in which the former has *tʰ- while the latter has *x-. The pronunciation of 天 in Middle Chinese as well as nearly all modern Sinitic varieties appears to originate from the *tʰ- variant. The *x- variant was later attested in Tang and Song texts, including Huìlín's 慧琳 *Yīqiè jīng yīnyì* 一切經音義 and *Jíyùn*, both specifying that the variant was found in Guānzhōng, part of the aforementioned Sīzhōu.[4] The western pronunciation of 天 seemingly indicates that the variety had been spoken since the Eastern Han dynasty and it was not confined to Guānzhōng in that era, i.e. it was used in the Central Plain as well. Nevertheless, it is unclear whether other western features can be traced back to Eastern Han; no direct attestation of them can be identified in the surviving materials. As a result, for the nonce, the beginning of Old Western Chinese is set at Sui/Tang in lieu of Eastern Han.

3. Association between Bai and Old Western Chinese

Some scholars have suggested that Bai and Old Western Chinese are closely related based on the pronunciations of 天 "sky" and 風 "wind" in the two lects. Lee and Sagart (2008: 373–4) associated the Bai word *xẽ55* "sky" with the aforementioned *x- variant of 天. In Baxter and Sagart's view (2014: 114), Bai *xẽ55* reflects a feature of an ancient western dialect where OC *lˤ- yields x-. A closely related case is 風 "wind" OC *prəm > MC *pjuwng* > *fēng*. Although the coda of 風 is -*ng* in Middle Chinese, this word rhymes with 心 "heart" MC *sim*, 南 "south" MC *nom*, and 欽 "admire" MC *khim*, in *Classic of Poetry* (*Shījīng* 詩經), all ending with -*m* in Middle Chinese; MC -*m* is a retention of OC *-m. On account of this, it is apparent that the coda of 風 is *-m in the first few centuries of the first millennium BCE.

---

[3] The Eastern Han forms are reconstructed by Baxter and Sagart (2014: 113).

[4] 祆, the phonetic radical of which is 天, is glossed as follows in Huìlín's *Yīqiè jīng yīnyì*: "祆神，上顯堅反，《考聲》云，胡謂神為天，今關中人謂天神為祆也。" [祆神 (Xiān deity), the first (word) is pronounced 顯堅反 (MC *xen*). The book *Kǎo shēng* says: the Iranians call deities 天 (heaven); nowadays people in Guānzhōng call heavenly deities 祆.] As for *Jíyùn*, 祆 is glossed as follows: "馨煙切⋯⋯關中謂天為祆。" [(pronounced as) *xen* ... people in Guānzhōng call sky 祆.] One can infer that 祆 is used to represent the Guānzhōng variant of 天.

Dialectal information of 風 can be found in the same chapter of *Shìmíng* in which 天 "sky" is glossed: 風 is pronounced as *piuŋ in Qīngzhōu and Xúzhōu, and as *pim in Yǎnzhōu, Yùzhōu, and Sīzhōu.[5] The variants of 風 have different codas: one ends in *-ŋ and the other in *-m. The Middle Chinese form *pjuwng* as well as those in virtually all modern varieties seemingly originate from the *-ŋ variant. Note that the geographical distributions of the two variants of 風 largely overlap with those of 天, and in both cases, Middle Chinese and mainstream modern varieties take the form used in Qīngzhōu and Xúzhōu. In Gong's view (2015: 10), the Bai word for "wind" *pi*[55] is seemingly akin to the *-m variant.[6] Indeed, 風 behaves differently from other *-(j)uwng* words in Bai as MC *-(j)uwng* typically corresponds to -ṽ in Bai.

As we shall see in Sections 3.1–3.4, there exist other commonalities between Bai and Old Western Chinese.

3.1 "Fungus"

In *Xīnjí zàngjīng yīnyì suíhán lù* 新集藏經音義隨函錄, written by Monk Kěhóng 可洪 in the tenth century CE, a dialectal word for "fungus" is attested:

(4) 卷十六：椹羹，上音審，菌生木上者也……正作蕈，山南土俗亦為審。按《五經字樣》作「蕈，式甚反」是也。[Volume 16: 椹羹, the first (character), pronounced 審 (MC *syim*[X]), refers to fungi on trees .... The correct character is 蕈 (MC *dzim*[X]), the word is also the vernacular term in Shānnán. Note: *Wǔjīng zìyàng* says "蕈, pronounced 式甚反 *syim*[X]".]

(5) 卷二十七：有蕈，音審，地菌也。應和尚亦音審。又按《字樣》作「式甚反」是也，漢上及蜀並呼菌為審也。[Volume 27: 有蕈, the second (character) pronounced 審 (MC *syim*[X]), refers to fungi on the ground. Monk Xuányìng also says it is pronounced 審. Note: *Zìyàng* says (its pronunciation is) 式甚反 *syim*[X], people in the upstream region of Hàn river and Sìchuān call fungus 審.]

Whether the dialectal word is a regional variant of 蕈 (just as 祆 is the regional variant of 天) or in the same word family with 蕈 remains to be seen; for now, the word is represented by 椹 in this study. It is apparent that it was used in western China and pronounced *syim*[X]. The Bai word for "fungus" *sɛ̃*[33] has been suggested to be 蕈 by Zhengzhang (1999: 52). This etymology is erroneous as *dz-* corresponds to ts-/tɕ- in Bai. *sɛ̃*[33] is most likely 椹: MC rising tone corresponds to Bai tone 33, *sy-* corresponds to Bai s-/ɕ-, and *-im* corresponds to Bai -ɛ̃ for words with *(t)sy-* in Middle Chinese.[7] Intriguingly, 椹 is also used in Caijia (*ɕaŋ*[55] "fungus"), a language which is closely akin to Bai.[8]

3.2 Words with zy- in Middle Chinese

It is stated in *Jíyùn* that 蛇 "snake" *zyae* and 射 "shoot" *zyae*[H] are pronounced *yae* and *yae*[H] in Guānzhōng; the reading *yae*[H] is used in 僕射 "executive assistant" *púyè*. Downer (1981: 8) noted that one variant of 船 "vessel" in *Jíyùn* – *ywen*, cf. standard *zywen*, is comparable to the examples above, though *Jíyùn* does not state where the variant was used. It appears

---

[5] The dialectal pronunciations are reconstructed by Gong (2015: 10).

[6] Nasalization is irregularly dropped in Jiànchuān, cf. Proto-Bai *pren1 (Wang 2006: 179).

[7] Compare 針 "needle" OC *t.kəm > MC *tsyim* > *zhēn* – Bai *tsɛ*[55], 枕 "pillow" OC *t.kəm² > MC *tsyim*[X] > *zhěn* – Bai *tsɛ*[33], 深 "deep" OC *l̥əm > MC *syim* > *shēn* – Bai *sɛ̃*[55].

[8] Middle Chinese rising tone corresponds to Caijia tone 55 (for words with a voiceless initial in MC), *sy-* corresponds to Caijia s-/ɕ-, *-im* corresponds to -aŋ.

that MC *zy-* corresponds to *y-* (presumably [j-]) in Old Western Chinese. In fact, this west-ern feature might date back to the Qin dynasty or even earlier, as the official title 僕射 was created in the Qin dynasty. In Bai, two *zy-* words have j-: ①食 "eat" OC *mə-lək > MC *zyik* > *shí* – Bai *ju⁴⁴*; ②船 "vessel" OC *Cə.lon > MC *zywen* > *chuán* – Bai *ji²¹*.[9]

3.3 Old Chinese *-p/t-s > -t

Haudricourt ([1954b](#)) proposed that Vietnamese became a tonal language through the loss of -ʔ and -h. In a similar fashion, he (Haudricourt [1954a](#)) hypothesized that the departing tone in Middle Chinese derives from earlier *-s. This hypothesis has been widely adapted into recent Old Chinese reconstructions. It is a well-known fact that whenever entering tone shows xiéshēng 諧聲 or etymological connections with another tone, that tone is nearly always the departing tone. As demonstrated by examples (6) to (8), both entering tone and departing tone are found in the same xiéshēng series. In example (6), the departing-tone word 內 "inside" is the phonetic radical of 納 "bring into", an entering-tone word. Moreover, there exist characters that have both entering- and departing-tone readings. When used as a verb, 度 is in the entering tone, but when it is a noun, it is in the departing tone (example (9)).

(6) 納 "bring into" OC *nˤup > MC *nop* > *nà*
　　內 "inside" OC *nˤup-s > MC *nwoj^H* > *nèi*
(7) 結 "tie (v.)" OC *kˤit > MC *ket* > *jié*
　　髻 "hair-knot" OC *kˤit-s > MC *kej^H* > *jì*
(8) 憶 "remember" OC *ʔrək > MC *'ik* > *yì*
　　意 "thought" OC *ʔrək-s > MC *'i^H* > *yì*
(9) 度 "measure (v.)" OC *dˤak > MC *dak* > *duó*
　　度 "measure (n.)" OC *dˤak-s > MC *du^H* > *dù*

Combining the *-s hypothesis and the xiéshēng contact between departing and entering tones, the pertinent departing-tone words are reconstructed as *-p/t/k-s in Old Chinese. For instance, in example (7), 髻 "hair-knot" is reconstructed as *kˤit-s, owing to its xiéshēng contact with 結 "tie (v.)". Baxter ([1992](#): 309) assumed the development of OC *-p/t/k-s up until Middle Chinese as shown below.

OC			MC
*-Vk-s	>*-Vs		>-V^H
*-Vt-s	>*-Vts	>*-Vjs	>-Vj^H
*-Vp-s	>*-Vts	>*-Vjs	>-Vj^H
*-Vwk-s	>*-Vws		>-Vw^H

　　A number of rhyme sequences in *Classic of Poetry* reflect that OC *-p-s had changed to *-t-s, e.g. Ode 257.13A, as shown in (10).[10] 對 "respond" is reconstructed as *tˤup-s in Old Chinese due to its probable etymological connection with synonymous 答 "answer" OC

---

[9] Alternatively, two words with *zy-* in Middle Chinese have an obstruent onset in Bai: 舌 "tongue" OC *mə.lat > MC *zyet* > *shé* – Bai *tse⁴²*, 舐 "lick" OC *Cə.leʔ > MC *zye^X* > *shì* – Bai *tsi³³*. The source of the distinction between j- and ts- requires further investigation.

[10] The labelling of rhyme sequences follows the convention of Baxter ([1992](#)).

*tˤup. The other four words are reconstructed as *-t-s. Evidence from the Chinese script also reflects this sound change. In the sense of "abandon", 廢 MC *pjoj^H* was originally written with 灋 (=法) OC *p.kap > MC *pjop* > *fǎ.* As 灋 ends in *-p, Baxter and Sagart (2014) reconstruct 廢 as *p-kap-s in lieu of *-t-s in Old Chinese. The character 廢 is of relatively late origin, and the use of phonetic radical 發 OC *Cə.pat > MC *pjot* > *fā* reflects the change of *-p-s to *-t-s. In the content beneath, Middle Old Chinese (MOC) *-t-s is used to represent the stage when OC *-p-s and *-t-s have merged, in order to disambiguate it from OC *-t-s.

> (10) 隧 "path" OC *sə-lut-s > MC *zwij^H* > *suì*
> 類 "good" OC *rut-s > MC *lwij^H* > *lèi*
> 對 "respond" OC *tˤup-s > MC *twoj^H* > *duì*
> 醉 "drunk" OC *Cə.tsut-s > MC *tswij^H* > *zuì*
> 悖 "silly" OC *bˤut-s > MC *bwoj^H* > *bèi*

As mentioned above, the sound change of *-p-s > MOC *-t-s may have taken place during the first half of the first millennium BCE as it is reflected in some rhyme sequences in *Classic of Poetry.* In a similar vein, there are signs that *-k-s became *-s in *Classic of Poetry.* For instance, *-ak-s often appears in the same rhyme sequence with *-a-s, e.g. Ode 166.1A, as seen in example (11). 庶 "numerous" ends in *-k-s whereas the other two words end in *-s. Meanwhile, there seems to be no tendency for *-p/t-s to be confused with *-j-s in the rhymes, indicating that the change of MOC *-t-s > *-js occurred later. Baxter and Sagart (2014: 196) implied that the change had taken place by the Han dynasty. Indeed, words with *-p/t-s in Old Chinese often rhyme with departing-tone words with OC *-j-s in Han rhymes, as shown in (12), reflecting that the stop in OC *-p/t-s was no longer present in that era.

> (11) 固 "firm" OC *kˤa-s > MC *ku^H* > *gù*
> 除 "give" OC *lra-s > MC *drjo^H* > *zhù*
> 庶 "numerous" OC *s-tak-s > MC *syo^H* > *shù*
> (12) Liú Xiàng 劉向 *Zhàng míng* 杖銘
> 味 "taste" OC *mət-s > MC *mj+j^H* > *wèi*
> 貴 "noble" OC *kuj-s > MC *kjw+j^H* > *guì*

The preface of the rhyme dictionary *Qièyùn* mentions several dialectal features in that era, including: 秦隴則去聲為入 "In Qín and Lǒng (roughly today's Shǎnxī and Gānsù), departing tone becomes entering tone". This indicates that (some) departing-tone words end in a stop. While *Qièyùn* and its revision *Guǎngyùn* 廣韻 do not provide any example of this regional feature, such examples are attested in some Tang and early Song works (Zhengzhang 2012: 57), as illustrated in examples (13) through (18). The form attested in *Guǎngyùn* (the Middle Chinese form), or the standard form prescribed in the pertinent text does not contain a stop coda, whereas that in Guānzhōng ends in *-t.* As I demonstrate below, words whose Old Chinese forms have been reconstructed end in *-p/t-s, and for those whose existence in Old Chinese is uncertain, their Middle Chinese form can also be projected back to *-p/t-s (in the case of (18), there is another possible source). In other words, the stop in the Middle Old Chinese cluster *-t-s is retained in Old Western Chinese.

> (13) 獪 "sly" MC *kwaj^H*/*kwaej^H* > *kuài*; 玄應《大唐眾經音義》卷十八：狡獪……古快反；《通俗文》：小兒戲謂之狡獪，今關中言狡刮，訛也。 [Xuányìng's *Dàtáng zhòngjīng yīnyì*, Volume 18: 狡獪... (獪) pronounced 古快反 *kwaej^H*; *Tōngsú wén* says

"small children playing" is called 狡獪 *kaew*[X] *kwaej*[H]; nowadays, people in Guānzhōng pronounce it as 狡刮 *kaew*[X] *kwaet*; this is an error.]

Xuányìng states that the standard pronunciation of 獪 is *kwaej*[H], whereas in Guānzhōng that is *kwaet*. Baxter and Sagart (2014) did not reconstruct the word's Old Chinese form. In their system, -*aej*[H] derives only from OC *-p/t-s. In Guānzhōng, the alveolar stop in MOC *-t-s is retained.[11]

(14) 蠆 "scorpion" MC *trhaej*[H] > *chài*; 玄應《大唐眾經音義》卷十八：蜂蠆，丑芥反，毒蟲也；山東呼為蠍，陝以西呼為蠆蝲，音土曷、力曷反。 [Xuányìng's *Dàtáng zhòngjīng yīnyì*, Volume 18: 蜂蠆, (蠆) pronounced 丑芥反 *trheaj*[H], is a venomous insect. People in Shāndōng call it 蠍, to the west of 陝(原) Shǎn(yuán), it is called 蠆蝲, pronounced 土曷(反)、力曷反 *that lat*.]

蠆 rhymes with 厲 "cruel" OC *rat-s > MC *ljej*[H] > *lì* and 邁 "proceed, march (v.)" OC *mˤrat-s > MC *maej*[H] > *mài* in *Classic of Poetry*. Its Old Chinese pronunciation is reconstructed as *mə-r̥ˤat-s.

(15) 稬 "chaff" MC *khwaj*[H]/*khwaej*[H] > *kuài*; 裴務齊正字本《刊謬補缺切韻》去聲泰韻：稬，苦會反，麁糖（糠），秦音苦活反。[Péi Wùqí revised *Kānmiù bǔquē qièyùn* 去聲 departing tone 泰韻 -*aj*: 稬, pronounced 苦會反 *khwaj*[H], means chaff; its Qín (Guānzhōng) pronunciation is *khwat*.]

The standard pronunciation of 稬 is stated as *khwaj*[H]. Baxter and Sagart (2014) did not reconstruct the word's Old Chinese form; in their system, -*aj*[H] comes only from OC *-p/t-s.

(16) 四 "four" MC *sij*[H] > *sì*;《集韻》息七切：四，關中謂四數爲四。[*Jíyùn* 息七切 *sit*: 四, "four" is pronounced *sit* in Guānzhōng.]

MC -*ij*[H] can be traced back to OC *-ij-s or *-i(p/t)-s. Baxter and Sagart (2014) reconstructed it as *s.li[j]-s, expressing uncertainty over the penultimate sound; 四 is better reconstructed as *-i[t]-s, partially due to the Guānzhōng variant *sit*; see Section 4.3 for further discussion.

(17) 淚 "tear (n.)" MC *lwij*[H] > *lèi*;《集韻》劣戌切：淚，關中謂目汁曰淚。[*Jíyùn* 劣戌切 *lwit*: 淚, people in Guānzhōng call "tear (n.)" *lwit*.]

On account of the semantic linkage between 泣 "weep" OC *k-r̥əp > MC *khip* > *qì* and 淚 "tear (n.)", 淚 is reconstructed as *rəp-s in Old Chinese.

(18) 顇 "distressed" MC *dzwij*[H] > *cuì*;《集韻》昨律切：顇……一曰關中謂癯弱為顦顇。[*Jíyùn* 昨律切 *dzwit*: 顇…. In Guānzhōng, "emaciation" is referred to as 顦顇 (顇: *dzwit*).]

Baxter and Sagart (2014) did not reconstruct the word's Old Chinese form. In their system, MC *dzwij*[H] comes from *-j-s or *-p/t-s. The Guānzhōng variant *dzwit* lends support to *-p/t-s.

---

[11] Although the phonetic radical of 獪 is 會 "meet, group" OC *m-kˤop-s, which is believed by many to be akin to 合 "join" OC *m-kˤop, it does not entail that 獪 has *-p-s in Old Chinese, since the character was likely created in the Han dynasty, which post-dates the change of OC *-p-s > MOC *-t-s. Also note that, according to the Western Han work *Fāngyán* 方言, 獪 was used in the regions of Qín 秦 and Jìn 晉 then.

Another potential example is the entering-tone reading of 鼻 "nose". According to *Guǎngyùn*, 鼻 is pronounced *bjij^H*, in departing tone, which is akin to the pronunciation of it in certain southern Chinese varieties, e.g. Yue and Hakka. Nonetheless, in Mandarin, Jin, and Wu, etc., the word is or reflects a form with a stop coda: Jianghuai Mandarin (Nánjīng) *piʔ⁵*, Wu (Shànghǎi) *biʔ¹*. The word is reconstructed as **m-bit-s* in Old Chinese. Pulleyblank (1973: 372) put forward that the entering-tone reading in varieties such as Jianghuai Mandarin was borrowed from the northwestern dialect.

The fate of OC **-k-s* in Old Western Chinese is less clear. Baxter (1992: 319) cited an example from Huìlín's *Yīqiè jīng yīnyì* in which **-k-s* seemingly became -k in Old Western Chinese:

> (19) 無復……下吳音扶救反，秦音馮目反。[無復 *wúfù*: .... The second (character) in the Wú 吳 pronunciation is 扶救反 *bjuw^H*; in the Qín 秦 pronunciation, it is 馮目反 *bjuwk*.]

In the sense of "again", 復 is reconstructed as **N-pruk-s* in Old Chinese (whence MC *bjuw^H*), and the Qín form can be analysed as preserving -k. Nevertheless, there is an alternative interpretation. In the sense of "return", 復 is reconstructed as OC **m-pruk* (whence MC *bjuwk*). The entering-tone reading may have displaced the departing-tone reading in Old Western Chinese, just as the level-tone reading of 釘 "nail (n.)" MC *teng > dīng* has superseded its departing-tone counterpart "nail (v.)" MC *teng^H > dìng* in Cantonese (both *tɛːŋ⁵⁵*). At the moment, there is inadequate evidence to conclude that OC **-k-s* developed into -k in Old Western Chinese.

I will now turn our attention to the case of Bai. Table 1 displays the tonal correspondences between Middle Chinese and Bai (Jiànchuān dialect). According to Starostin (1995: 175), departing tone corresponds to either tone 31 or 42 in Bai. Zhengzhang (1999: 20–1) delivered a more precise analysis: departing tone corresponds to tone 42 in Bai for words with a voiceless initial in Middle Chinese and to tone 31 for those with a voiced initial. Nonetheless, his analysis is still partially spurious; in my view, the correct analysis is as follows: words with a voiceless aspirated or voiced obstruent initial in Middle Chinese have tone 31, whereas words with a sonorant or voiceless unaspirated initial (including ʔ-) have tone 42.[12] Tone split based on aspiration is rare in Sinitic languages.[13] Other than this feature, there exists another captivating feature regarding departing-tone words in Bai. First noted by Starostin (1995: 175–6), a set of departing-tone words, termed EL (entering-tone-like) words in the present study, behave as if they were entering-tone words in terms of tone. Entering-tone words (which no longer retain the stop coda in Bai) with a voiceless or sonorant initial in Middle Chinese are in tone 44 in Bai, while those with a voiced obstruent initial are in tone 42. Likewise, EL words with a voiced obstruent initial in Middle Chinese are in tone 42, while other EL words are in tone 44; EL words are shown in Table 2. In addition to the EL words Starostin identified, there are three more EL words, namely 氣 "air", 醉 "drunk", and 彘 "pig".

Starostin posited that the EL words he identified contain a front stop consonant (coda), which is probably followed by **-s*, in Old Chinese; this cluster merged with **-j* in mainstream Chinese no later than the fourth century ᴄᴇ. As can be seen in Table 2, most EL

---

[12] In the case of MC *s-* (see Table 3), the distinction between tone 31 (e.g. 細 "small") and 42 (算 "calculate") cannot be accounted for in terms of Middle Chinese. In some Bai dialects, e.g. Hèqìng 鶴慶 dialect, there exists a distinction between aspirated and unaspirated fricatives (Zhao 2010: 27). *s-* words with tone 31 (in Jiànchuān) in Table 3 are pronounced *sʰ/ɕʰ-*, whereas 算 is pronounced *s-*. In other words, the tonal distinction is based on aspiration in earlier stages of Bai, which is not retained in the Jiànchuān dialect. Whether this aspiration distinction in fricatives can be traced back to Old Chinese remains to be seen.

[13] One example is the Xiāngxiāng 湘鄉 dialect of Xiang (Chang 1975). Intriguingly, in such dialects, the tone is lowered in the presence of aspiration (Chang 1975: 673–4; Sagart 1981).

Table 1. Tonal correspondences between Middle Chinese and Bai

MC tone		Bai tone			
		Nature of MC initial			
		Voiceless		Sonorant	Voiced obstruent
		ASP	UNASP		
Level		55			21
Rising		33			
Departing	Non-EL	31	42		31
	EL	44			42
Entering		44			42

words, or nine out of eleven EL words, end with *-p/t-s in Baxter-Sagart's Old Chinese reconstruction. In the case of 四 "four" OC *s.lij-s (or *s.li[j]-s) > MC *sij^H*, the fact that it is in the EL group, along with other evidence, lends support to reconstructing it as *-i(p/t)-s; see Section 4.3 for further discussion.

It is then necessary to examine whether *-p/t-s is absent in non-EL words. Starostin listed only EL words but not non-EL words, and thus it is unclear how many non-EL words he studied. The present study has examined 55 non-EL words, listed in Table 3. Of the 47 words whose OC form is available, only one contains *-t-s: 利 "sharp" OC *C.rit-s > MC *lij^H* > *lì* – Bai *ji^31*.[14] The Old Chinese forms of words 鋪 "store (n.)" MC *phu^H*, 糩 "fart" MC *phj+j^H*, 踞 "[Bai] sit" *kjo^H*, 餵 "feed" *ʼjwe^H*, 轉 "revolve, turn" *trjwen^H*, 硬 "hard" *ngeang^H*, 瞑 "[Bai] dark" *meng^H*, and 麪 "noodles" *men^H* have not been reconstructed.[15] In the Baxter-Sagart system, 糩 *phj+j^H* is the only word that may derive from *-p/t-s, given that it did exist in Old Chinese. In all likelihood, 糩 *phj+j^H* is a variant of 屁 "fart" *phjij^H*; their Middle Chinese forms come from either OC *-j-s or *-p/t-s. Both characters are of late origin and were likely created after the change of MOC *-t-s > *-js, and hence the choice of phonetic radicals is not indicative of their Old Chinese form.[16] In all honesty, the word(s) 糩/屁 probably did not exist in Old Chinese in the first place.

That OC *-p/t-s is commonplace in the EL group but practically absent in the non-EL group is by no means a coincidence. Since *-p/t-s has the same tonal reflexes as OC *-p/t/k in Bai, *-p/t-s probably lost *-s and became a stop coda sometime in the past; this change most likely post-dates OC *-p/t-s > MOC *-t-s. As illustrated above, the changes of OC *-p-s > *-t-s and OC *-k-s > *-s are both observed in multiple rhyme sequences in *Classic of Poetry*. The latter change is reflected in Bai, as words such as 刺 "thorn" OC *tsʰek-s, 細 "small" OC *sˤek-s, 戴 "wear" OC *Cˤə.tˤək-s, 箸 "chopsticks" OC *drak-s, etc. are in the non-EL group, revealing that the stop in *-k-s was not retained in early stages of Bai. As early Bai had undergone this change, one can deduce that it had also undergone OC *-p-s > *-t-s (i.e. joining OC *-t-s) when EL words lost *-s in the cluster. It is then reasonable to assume that EL words subsequently ended in *-t.

---

[14] One interpretation is that it has developed an irregular tone; another interpretation is that the tone implies that it ends in *-j-s in Old Chinese.

[15] [Bai] indicates the meaning is applicable to Bai, but not mainstream modern Chinese varieties.

[16] In point of fact, their phonetic radicals point to different Old Chinese endings, compare 費 OC *pʰut-s vs 庇 OC *pij-s.

Table 2. EL words

MC initial – Bai tone	Word	OC	MC	Mandarin	Bai
MC voiceless initial – Bai tone 44	四 "four"	*s.lij-s	sij^H	sì	ɕi⁴⁴
	肺 "lung"	*pʰot-s	$phjoj^H$	fèi	pʰiɑ⁴⁴
	氣 "air"	*C.qʰəp-s	$khj+j^H$	qì	tɕʰi⁴⁴
	歲 "year, age"	*s-qʷʰat-s	$sjwej^H$	suì	suɑ⁴⁴
	醉 "drunk"	*Cə.tsut-s	$tswij^H$	zuì	tɕui⁴⁴
MC sonorant initial – Bai tone 44	二 "two"	*nij-s¹⁷	$nyij^H$	èr	ne⁴⁴
	外 "outside"	*ŋʷˤat-s	$ngwaj^H$	wài	ŋuɑ⁴⁴
MC voiced obstruent initial – Bai tone 42	大 "big"	*lˤat-s	$da(j)^H$	dà	to⁴²
	吠 "bark (v.)"	*Cə.bot-s	$bjoj^H$	fèi	piɑ⁴²
	胃 "stomach"	*ɢʷət-s	$hjw+j^H$	wèi	vu⁴²
	彘 "pig"	*lrat-s	$drjej^H$	zhì	te⁴²

3.4 Words with d- and rising tone in Middle Chinese

Two words with *d-* in Middle Chinese are pronounced *th-* in Old Western Chinese, namely 稻 "paddy" OC *lˤuʔ > MC *daw^X* > *dào* and 墮 "fall (v.)" OC *lˤojʔ > MC *dwa^X* > *duò*. Examples (20) to (22), cited from Tang and Song works, demonstrate that 稻 was pronounced *thaw^X* in Guānzhōng. Examples (23) and (24), which are comments (pertinent) to a line of a poem by the Tang poet Dù Fǔ 杜甫, reveal that 墮 was pronounced *thwa^X* in that region.

(20) 《國史補》：關中人呼稻為討。[*Guóshǐ bǔ*: People in Guānzhōng pronounce 稻 as 討 *thaw^X*.]

(21) 《集韻》土晧切：稻，秔也，關西語。[*Jíyùn* 土晧切 *thaw^X*: 稻, non-glutinous rice; this pronunciation is found in Guānxī.]

(22) 《集韻》土晧切：稄，關西呼蜀黍曰稄黍。[*Jíyùn* 土晧切 *thaw^X*: 稄, in Guānxī, sorghum is called 稄黍.]

(23) 《邵氏聞見後錄》：潘邠老云「花妥鶯捎蝶，溪喧獺趁魚。妥音墮，乃韻。」邠老不知秦音以落為妥……少陵，秦人也。[*Shàoshì wénjiàn hòulù*: Pān Bīnlǎo says "花妥鶯捎蝶，溪喧獺趁魚 huā tuǒ yīng shāo dié, xī xuān tǎ chèn yú, 妥 *thwa^X* is pronounced 墮…". Pān is not aware that "fall (v.)" is pronounced 妥 in Qín … 少陵 Shàolíng (杜甫 Dù Fǔ) is from Qín.]¹⁸

(24) 《苕溪漁隱叢話》：西北方言以墮為妥，花妥即花墮也。[*Tiáoxī yúyǐn cónghuà*: 墮 becomes 妥 in the northwestern dialect, 花妥 is actually 花墮.]

This is a comment to the line 花妥鶯捎蝶 mentioned in example (23).

---

¹⁷ The Bai form of 二 can be regarded as evidence supporting *-p/t-s. Nonetheless, there is no additional evidence from the Chinese script, rhyming in *Classic of Poetry*, or living Sinitic varieties.

¹⁸ The birthplace of Dù Fǔ is disputed; it is claimed to be Gǒngyì 鞏義 by some and Luòyáng 洛陽 by others. Both places are not situated in Guānzhōng, so it is possible that the comment "少陵，秦人也" [Shàolíng is from Qín] does not mean the poet was born in Guānzhōng. The poem that contains "花妥鶯捎蝶，溪喧獺趁魚" is known to have been written when Dù Fǔ lived in Xī'ān 西安 (then Cháng'ān 長安), and hence 秦人 (a person from Qín) may refer to where he resided then.

Table 3. Non-EL words

MC initial – Bai tone	Word	OC	MC	Mandarin	Bai
MC voiceless aspirated initial – Bai tone 31	寸 "inch"	*tshˤun-s	tshwon^H	cùn	tɕʰuĩ^{31}
	片 "(classifier)"	*pʰˤen-s	phen^H	piàn	pʰĩ^{31}
	刺 "thorn"	*tsʰek-s	tshje^H	cì	tɕʰi^{31}
	信 "believe"	*s-niŋ-s	sin^H	xìn	ɕĩ^{31}
	炭 "charcoal"	*tʰˤan-s	than^H	tàn	tʰã^{31}
	破 "break, chop"	*pʰˤaj-s	pha^H	pò	pʰo^{31}
	細 "small"	*sˤek-s	sej^H	xì	se^{31}
	處 "place"	*t.qʰaʔ-s	tsyho^H	chù	tsʰʮ^{31}
	菜 "vegetables"	*s.r̥ˤəʔ-s	tshoj^H	cài	tsʰɯ^{31}
	蒜 "garlic"	*sˤor-s	swan^H	suàn	suã^{31}
	鋪 "store (n.)"	–	phu^H	pù	pʰu^{31}
	糞 "fart"	–	phj+j^H	fèi	fʮ^{31}
	覆 "lid, cover"	*pʰruk-s	phjuw^H	fù	pʰɯ^{31}
MC voiceless unaspirated initial – Bai tone 42	半 "half"	*pˤan-s	pan^H	bàn	pã^{42}
	正 "[Bai] right (≠left)"[19]	*teŋ-s	tsyeng^H	zhèng	tsɛ^{42}
	衣 "wear"	*ʔrəj-s	ʔj+j^H	yì	ji^{42}
	見 "see"	*kˤen-s	ken^H	jiàn	kẽ^{42}
	咽 "swallow (v.)"	*ʔˤin-s	ʔen^H	yàn	ẽ^{42}
	豹 "leopard"	*pˤrewk-s	paew^H	bào	pã^{42}
	釘 "nail (v.)"	*tˤeŋ-s	teng^H	dìng	tɕẽ^{42}
	救 "save (v.)"	*s.kru-s	kjuw^H	jiù	kuu^{42}
	過 "pass (v.)"	*kʷˤaj-s	kwa^H	guò	ko^{42}
	種 "plant (v.)"	*mə-toŋʔ-s	tsyowng^H	zhòng	tsʮ^{42}
	算 "calculate"	*sˤorʔ-s	swan^H	suàn	suã^{42}
	價 "price"	*C.qˤraʔ-s	kae^H	jià	kɛ^{42}
	箭 "arrow"	*tsen-s	tsjen^H	jiàn	tɕĩ^{42}
	踞 "[Bai] sit"	–	kjo^H	jù	kʮ^{42}
	戴 "wear"	*Cə.tˤək-s	toj^H	dài	tũ^{42}
	濟 "cross a river"	*tsˤijʔ-s	tsej^H	jì	tɕi^{42}
	餵 "feed"	–	ʔjwe^H	wèi	ui^{42}

(Continued)

---

[19] Nasalization is irregularly dropped in Jiànchuān Bai, cf. Hèqìng Bai tsə̰r^{42} (Zhao 2010: 246). Also note that "right (side)" is also represented by 正 in some Min 閩 varieties.

Table 3. (*Continued.*)

MC initial – Bai tone	Word	OC	MC	Mandarin	Bai
	轉 "revolve, turn"	–	$trjwen^H$	zhuàn	tsuĩ⁴²
	醬 "sauce"	*tsaŋ-s	$tsjang^H$	jiàng	tɕõ⁴²
	鏡 "mirror"	*C.qraŋʔ-s	$kjaeng^H$	jìng	kɛ̃⁴²
	變 "change"	*pron-s	$pjen^H$	biàn	pĩ⁴²
MC sonorant initial – Bai tone 42	命 "life"	*m-riŋ-s	$mjaeng^H$	mìng	miɛ⁴²
	面 "side, face"	*C.men-s	$mjien^H$	miàn	mi⁴²
	硬 "hard"	–	$ngeang^H$	yìng	ŋɛ⁴²
	夢 "dream"	*C.məŋ-s	$mjuwng^H$	mèng	(mə̃⁴²)[20]
	暝 "[Bai] dark"	–	$meng^H$	míng	miɛ⁴²
	蔓 "vine"	*C.man-s	$mjon^H$	wàn/màn	me⁴²
	餓 "hungry"	*ŋˤaj-s	nga^H	è	ŋo⁴²
	麪 "noodles"	–	men^H	miàn	mi⁴²
MC voiced obstruent initial – Bai tone 31	地 "land, earth"	*lˤej-s	dij^H	dì	tɕi³¹
	步 "step"	*mə-bˤa-s	bu^H	bù	pu³¹
	豆 "bean"	*N.tˤo-s	duw^H	dòu	tɯ³¹
	病 "ill"	*braŋ-s	$bjaeng^H$	bìng	pɛ̃³¹
	淨 "[Bai] shave"	*m-tseŋ-s	$dzjeng^H$	jìng	tɕɛ̃³¹
	剩 "remain"	*Cə.ləŋ-s	$zying^H$	shèng	sũ³¹
	盜 "steal"	*dˤawk-s	daw^H	dào	tɑ³¹
	箸 "chopsticks"	*drak-s	$drjo^H$	zhù	tsɣ³¹
	樹 "tree"	*m-toʔ-s	$dzyu^H$	shù	tsɯ³¹
	櫃 "cupboard"	*gruj-s	$gwij^H$	guì	kɣ³¹
	舊 "old"	*N-kʷəʔ-s	$gjuw^H$	jiù	ku³¹
	(利 "sharp")[21]	*C.rit-s	lij^H	lì	ji³¹
	(漏 "leak (v.)")[22]	*Nə-rˤok-s	luw^H	lòu	ɣu³¹

On the surface, these two Guānzhōng forms seemingly indicate that voiced stops (or affricates) were changed to voiceless aspirated ones in the lect. Nevertheless, this phenomenon is not mirrored in other attested Guānzhōng variants of words with voiced stops/

[20] According to Wang (2006), the Proto-Bai form of 夢 is *muu⁵. Proto-Bai tone 5 yields tone 42 in Jiànchuān dialect. Thus, the tone of the Jiànchuān form *muu³¹* is irregular. The form cited in the table is from Dàshí 大石 dialect, where Proto-Bai tone 5 also yields tone 42.

[21] Words with *l-* in Middle Chinese have ɣ-/j- or k- in Bai, depending on their Old Chinese onset (Gong 2015: 9). The Bai finals with which ɣ- and j- are combined are in complementary distribution: j- is combined with -i/ĩ and ɣ- with other finals (e.g. 漏 "leak (v.)"). As such, for those with j-, their onset may have been *ɣ- in early stages of Bai, and thus the word 利 is placed here.

[22] See note 21.

affricates in Middle Chinese in that era. In examples (25) through (29), while the initials (stops/affricates) of the Middle Chinese form and the Guānzhōng form may not share the same place of articulation, they are both voiced. Also worthy of note, the Middle Chinese initial of the word 糴 "buy (grain)" in example (29) is *d-*, but the initial of its Guānzhōng form is different from that of 稻 and 墮. One of the differences between them lies in tone: 糴 is in entering tone, whereas the other two words are in rising tone.

(25) 尚 "in charge of" MC *dzyang* > *shàng*; 《夢溪筆談》補筆談卷一: 官名中尚書，本秦官，尚音上……至今秦人謂尚為常。[*Dream Pool Essays* 補筆談 *bǔbǐ tán* Volume 1: Among the titles of officials, 尚書 has its origin in Qín; 尚 is pronounced 上 *dzyang*[H] .... Up to now, people in Qín still pronounce 尚 as 常 *dzyang*.]

According to *Guǎngyùn*, the pronunciation of 尚 is *dzyang* for 尚書.

(26) 蟲 "insect" MC *drjuwng* > *chóng*; 《中山詩話》 : 周人語轉，亦如關中以中為蒸，蟲為塵，丹青之青為蔶也。[*Zhōngshān shīhuà*: The Zhōu people changed the pronunciation, just as in Guānzhōng, 中 *trjuwng* is pronounced 蒸 *tsying*, 蟲 *drjuwng* is pronounced 塵 *drin*, and 青 *tsheng*, as in 丹青, is pronounced 蔶 *tshej*.]

(27) 晨 "dawn" MC *dzyin* > *chén*; 《集韻》慈鄰切 : 晨，旦也，關中語。[*Jíyùn* 慈鄰切 *dzin*: 晨, "dawn"; this pronunciation is used in Guānzhōng.]

(28) 顇 "distressed" MC *dzwij*[H] > *cuì*; 《集韻》昨律切 : 顇……一曰關中謂瘦弱為顦顇。[*Jíyùn* 昨律切 *dzwit*: 顇 .... In Guānzhōng, "emaciation" is referred to as 顦顇 (顇: *dzwit*).]

(29) 糴 "buy (grain)" MC *dek* > *dí*; 《集韻》直畧切 : 糴，關中謂買粟麥曰糴。[*Jíyùn* 直畧切 *drjak*: 糴, in Guānzhōng, buying grain is called 糴.]

Intriguingly, similar patterns can be discerned in Bai. Words with voiced stops/affricates in Middle Chinese have a voiceless unaspirated onset in Bai, and exceptions are rare. In the case of MC *d-*, it corresponds to t- or tɕ- (through secondary palatalization) in Bai. Words with MC *d-* that are found in Bai are shown in Table 4. Two words have a voiceless aspirated onset in Bai, namely 弟 "younger brother" *tʰi³³* and 道 "road" *tʰu³³*; both words are in rising tone (from OC *-ʔ) in Middle Chinese. Rising-tone words with other Middle Chinese stops/affricates do not have an aspirated onset in Bai, e.g. 庳 "low, short (stature)" MC *bjie*[X] – Bai *pi³³*, 重 "heavy" MC *drjowng*[X] – Bai *tsɣ̃³³*, 舅 "maternal uncle" MC *gjuw*[X] – Bai *ku³³*. Note that 待 "wait" in Table 4 is also in rising tone but its onset is unaspirated; it has *dˤ- in Old Chinese while 弟 and 道 have *lˤ-. As such, it appears that 弟 and 道 are distinct from other *d-* words because they have both *lˤ- and *-ʔ in Old Chinese. Parallels can be drawn from Caijia and Longjia, languages that have been claimed to be closely akin to Bai (Zhengzhang 2010; Lee 2023). The onset of *lˤ- words in Caijia and Longjia is either l- or a fricative (x-/h-/ɬ-), as shown in Table 5.[23] The words with a fricative onset happen to be 弟 and 道.

As can be seen, words with OC *lˤ- and *-ʔ do not have the same onset as other *d-* words in Bai, Caijia, and Longjia. In the case of Old Western Chinese, 稻 "paddy" and 墮 "fall (v.)" also have *lˤ- and *-ʔ in Old Chinese; we are unable to compare them with *d-* words with rising tone and T-type onset in OC (e.g. 待 "wait") since the latter are not attested, but it is possible that a distinction in onset equivalent to the one observed in Bai existed. A question that follows is the mechanism that made words with *lˤ- and *-ʔ distinct. *lˤ-, a pharyngealized consonant and *-ʔ, a glottal stop, are both laryngeal elements. It is possible that *lˤ- underwent dissimilation due to the

---

[23] The distinction between the Old Chinese sources of MC *d-*, namely T-type (e.g. *dˤ-, *m-tˤ-) and L type (*lˤ-) onsets is preserved in Caijia and Longjia, as illustrated in Table 5.

Table 4. Words with *d-* in Middle Chinese found in Bai

MC tone	Word	OC	MC	Mandarin	Bai
Level tone	桃 "peach"	*C.lˤaw	*daw*	táo	tɑ²¹
	啼 "[Bai] sing"	*C.lˤe	*dej*	tí	tɕi²¹
	填 "fill up"	*dˤin	*den*	tián	tɕi²¹
	痰 "phlegm"	–	*dam*	tán	tɑ̃²¹
	銅 "copper"	*lˤoŋ	*duwng*	tóng	tɣ̃²¹
	頭 "head"	*m-tˤo	*duw*	tóu	tu²¹
Rising tone	弟 "younger brother"	*lˤəjʔ	*dej*^{X}	dì	tʰi³³
	道 "road"	*kə.lˤuʔ	*daw*^{X}	dào	tʰu³³
	待 "wait"	*dˤəʔ	*doj*^{X}	dài	tu³³
Departing Tone	大 "big"	*lˤat-s	*da(j)*^{H}	dà	to⁴²
	地 "land, earth"	*lˤej-s	*dij*^{H}	dì	tɕi³¹
	豆 "bean"	*N.tˤo-s	*duw*^{H}	dòu	tu³¹
	盜 "steal"	*dˤawk-s	*daw*^{H}	dào	tɑ³¹
Entering tone	蹋 "tread on"	*lˤap	*dap*	tà	tɑ⁴²

Table 5. *d-* words found in various languages

Word	OC	MC	Mandarin	Longjia	Caijia	Bai
豆 "bean"	*N.tˤo-s	*duw*^{H}	dòu	ntau⁵⁵	tu²²	tu³¹
頭 "head"	*m-tˤo	*duw*	tóu	tau⁵⁵	tu²¹	tu²¹
大 "big"	*lˤat-s	*da(j)*^{H}	dà	la⁵⁵	la²²	to⁴²
田 "field"	*lˤiŋ	*den*	tián	li⁵⁵	ləŋ²¹	—
地 "land, earth"	*lˤej-s	*dij*^{H}	dì	li³³	le²²	tɕi³¹
弟 "younger brother"	*lˤəjʔ	*dej*^{X}	dì	ɬɛ³³	he⁵⁵	tʰi³³
道 "road"	*kə.lˤuʔ	*daw*^{X}	dào	hau³¹	xo⁵⁵	tʰu³³

presence of *-ʔ, though the phonetic value of the resulting sound is unclear. Note that in case this hypothesis is sound, it may imply that by the time the change occurred in the lect that gave rise to Old Western Chinese and Bai, *lˤ- might be the only onset that retained pharyngealization, otherwise we would expect words with other pharyngealized onsets and *-ʔ, e.g. 抱 "hug" OC *m-pˤuʔ – Bai *pu³³*, to be distinct as well. More work needs to be done to see if there is additional evidence to substantiate the above hypothesis.

4. Discussion

4.1 Northwest Chinese in Medieval China as seen from Coblin's studies

In examining the features of Old Western Chinese, the present study rests upon direct attestations of dialectal variants in Chinese texts. Other scholars focus on transcriptional

Table 6. Stages of Northwest Chinese in Medieval China

Pre-Old Northwest Chinese (Pre-ONWC)	Dm: Dharmarakṣa	
Old Northwest Chinese (ONWC)	Cháng'ān dialect	Km: Kumārajīva
	Corridor dialect	BZ: Buddhayaśas and Zhú Fóniàn (the underlying Indic language is likely Gāndhārī (Gd.)) Dk: Dharmakṣema
Sui-Tang Cháng'ān (STCA)	Jn: Jñānagupta YSG: Yán Shīgǔ	
Mid-Tang Cháng'ān (MTCA)	Am: Amoghavajra HL: Huìlín	
Late Tang Cháng'ān (LTCA)	S-T: Sino-Tibetan Treaty Inscription of 821–22	
Common Shāzhōu (CSZ)	C: *Qiānzìwén* 千字文 K: *Jīngāngjīng* 金剛經 O: *Ēmítuójīng* 阿彌陀經 T: *Dàshèng zhōngzōng jiànjiě* 大乘中宗見解 TD: *Tiāndì bāyáng shénzhòujīng* 天地八陽神咒經 NT: *Nántiānzhú guó pútídámó chánshī guānmén* 南天竺國菩提達磨禪師觀門 DA: *Dào'ān fǎshī niànfó zàn* 道安法師念佛讚 Kbr: Khotanese Brāhmī materials	
Colloquial Shāzhōu (CollSZ)	T: Tibetan forms	

materials and fǎnqiè 反切 glosses in commentaries on Buddhist texts or Chinese classic texts in their endeavours to unveil ancient western dialects, for instance, Coblin (1991, 1992, 1994a, 1994b); he studied the phonology of ancient northwestern dialects with the help of such materials. The subjects of his studies are the historical varieties spoken in Guānzhōng and the Gānsù Corridor in Medieval China (see Table 6). The earliest stage is called Pre-Old Northwest Chinese (Pre-ONWC), dated around 280 CE, and the reconstruction of which is based upon transcriptional data from the corpus of Dharmarakṣa; he was active both at Cháng'ān and at various places in the Gānsù Corridor, and Coblin did not specify the nature of the underlying Chinese variety. The usable material for reconstruction is altogether rather scant, and thus only a modicum of words has Pre-ONWC forms in Coblin (1994a). The next stage is Old Northwest Chinese (ONWC), dated to about 400 CE. In Coblin's analysis, there exist two varieties, namely Cháng'ān dialect and Corridor dialect. Kumārajīva's transcriptional corpus is thought to represent the former, whereas those of Buddhayaśas and Zhú Fóniàn 竺佛念 and Dharmakṣema represent the Corridor dialect. Sui-Tang Cháng'ān (STCA) represents the dialect of Cháng'ān during the Sui and early Tang period; the reconstruction of it is based on Jñānagupta's Buddhist transcription, as well as the fǎnqiè glosses and direct sound annotations in Yán Shīgǔ's 顏師古 commentary to the *Hànshū* 漢書. Mid-Tang Cháng'ān (MTCA) refers to the Cháng'ān dialect spoken in the eighth century; the sources for

reconstructing it are the transcriptional corpus of Amoghavajra and the fǎnqiè/direct sound glosses in Huìlín's *Yīqiè jīng yīnyì*. The subsequent stage is Late Tang Cháng'ān (LTCA); its reconstruction is based upon the Sino-Tibetan Treaty Inscription of 821–22. The sixth stage is Common Shāzhōu (CSZ). A number of medieval northwest dialects are reflected in Tibetan and, to a lesser extent, in Brāhmī transcriptional materials from Dūnhuáng, which are dated to the ninth and tenth centuries. The dialect forms can be compared to yield Common Shāzhōu reconstructions, representing the common ancestor of the Shāzhōu (SZ) dialects. The stage of Colloquial Shāzhōu (CollSZ) refers to actual examples of the colloquial speech of the Shāzhōu area preserved in the Tibeto-Chinese and Brāhmī-Chinese phrase books. Several sources of the CSZ stage are excluded from the present study, as the pertinent data does not bear on the discussion beneath.

As opposed to direct attestations, a significantly larger number of Chinese words are employed in the transcriptional materials and glosses in commentaries, and thus the latter provides a more thorough picture of the underlying Chinese varieties. Nonetheless, several issues need to be kept in mind when using these materials.

First, the identities of the underlying varieties are not always clear. For instance, in discussing the Chinese side of Kumārajīva's corpus, Coblin (1991: 8) stated that, "We may guess that the variety underlying the transcriptions was in the main northwestern, but we cannot rule out the possibility of internal inconsistency where teamwork of the sort envisaged here was involved." Likewise, judging from the statistical tendency to echo the *Qièyùn* distinctions in Huìlín's *Yīqiè jīng yīnyì*, Coblin (1994a: 23) commented that, "HL's [Huìlín] work may indeed be based on an older, more finely drawn canvas of sources or traditions upon which later, broader strokes of current pronunciation have been superimposed."

Furthermore, for some materials, it is difficult to sift out all the forms from earlier periods. Concerning Kumārajīva's corpus, Coblin (1991: 9) wrote, "There remains, therefore, the unfortunate possibility that material from these sources (Wèi-Jìn period texts) has been adopted by Km's [Kumārajīva] translation team and has consequently been falsely included in our study."

Notwithstanding the limitations, it is worth examining whether the features discussed in Section 3 are present in the transcriptional materials and glosses to see if the direct attestations are in line with these materials. In the content beneath, the abbreviations, reconstructed forms, Indic forms (P.=Pali, Skt.=Sanskrit), and other transcriptional forms are cited from Coblin (1994a).

天 **"sky"**

In Buddhayaśas and Zhú Fóniàn's transcriptional materials, Chinese 天 renders Indic *h*-, e.g. 摩天提伽 for Sanskrit *maharddhika* (equivalent to P. *mahiddhika*/ Gd. *mahedhiǵa~mahedhiya*) (Coblin 1994b: 155–6) (see Table 7). The word is employed in multiple materials dated the ninth/tenth centuries CE in which the underlying Chinese varieties are believed to be Shāzhōu dialects, and the word renders *th*- in place of *h*-. In Huìlín's *Yīqiè jīng yīnyì*, 天 constantly acts as the upper speller for words with *th*- in MC, and thus, the reading represented by this character is pronounced *th*- as well. The *x*- variant is represented by 祆; see footnote 4 for details.

Words with zy- in Middle Chinese

It has been suggested that 蛇維, as found in Dharmarakṣa's work, is used to transcribe a Gāndhārī form equivalent to Skt./P. *jhāpita*; its phonetic form is probably *zavi/*źavi (see Brough (1962: 59–62)) (see Table 8). In Buddhayaśas and Zhú Fóniàn's work, 蛇婆提伽

Table 7. 天 "sky"

ONWC	STCA	MTCA	CSZ/CollSZ
*ḣen~*ṫhen BZ: -har(d)- (=Gd. *-he(d)-, P. -hid-)	*thian	*thian > *thian	CSZ *thian K, O, TD, NT: then; DA: thyen CollSZ *thian T: then

Table 8. 蛇 "snake"

Word	Pre-ONWC	ONWC	STCA	MTCA	CSZ
蛇 OC *Cə.lAj MC *zyae*	*ja or *źa (?) Dm: jhā- (perhaps =Gd. *zavi or *źavi?)	*ia, (~*źa?) BZ: ya-, P. -kā/ Skt. kaḥ (=Gd. *-ya)	*ia, (~*źa?) Jn: -yā-	*źa HL: 常耶, 社耶, 食遮, 射遮, 時遮, 常遮, 社遮	*śa T: sha

renders P. *yamaṭaggi*; another example is 帶叉蛇婆提, in which 蛇 renders Gd. *-y-*. (蛇 corresponds to P. *-kā*/Skt. *-kaḥ*, and Gāndhārī intervocalic *-k-* is often reduced to *-y-*.) This same "ya-reading" for 蛇 appears in the transcription of Jñānagupta as well: 毘梨蛇耶 renders Skt. *vīryāya*. Nevertheless, 蛇 represents *ź-* and *ś-* in MTCA and CSZ respectively. Note that words with *y-* in MC do not have *ź-* in MTCA and *ś-* in CSZ: 耶 MC *yae* MTCA/CSZ *ia*.

With regard to 射, there exist four readings in *Guǎngyùn*: *zyae^H*, *zyek*, *yae^H*, *yek*. As mentioned in Section 3.2, the third reading is found in 僕射. In Coblin (1994a), there is only one entry for 射 (*yae^H*): STCA *ia YSG 弋舍 (=*yae^H*); LTCA *ia S-T: *ya*. In the Sino-Tibetan Treaty Inscription, *ya* is exactly used to render 僕射; it is uncertain whether 射 is also pronounced *ya* in other cases in the Chinese variety. As for Yán Shīgǔ's commentary, in fact, *zyek* and *yek* can also be found; in addition, the reading *yae^H* is only used in the names of species, and it might not be the standard reading in the commentary's underlying variety.

OC *-p/t-s > *-C

Several words with *-p/t-s in Old Chinese seemingly end in an obstruent coda in Pre-ONWC/ONWC; for the ONWC period, they are found in works in which the underlying Chinese variety is presumably the Corridor dialect but not in Kumārajīva's transcription where the Chinese side is likely the Guānzhōng dialect. As shown in Table 9, 帶, 害, 賴, and 會, all having *-aj^H* in MC, are employed to transcribe Indic syllables ending in an obstruent coda. Coblin (1991: 70–73) reconstructs them, as well as other words with *-aj^H*, as *-aC and *-ai in the Corridor and Guānzhōng dialects respectively. In the case of the Corridor dialect, he put forward that the "unassimilated" value of the sound represented by the cover symbol *-C is *-ś*, which is found before a bilabial nasal, whereas the sound is realized as *-t* before oral and nasal dental stops. In the MTCA and CSZ (or CollSZ) periods, the abovementioned four words no longer contain an obstruent coda.

For the STCA variety, the reflex of OC *-p/t-s does not contain a stop coda, save the suspicious case of 四. In Jñānagupta's transcription, Sanskrit *siddhyantu* is transcribed as 四填妒 (MC *sij^H den tu^H*) or 膝填妒 (MC *sit den tu^H*). The second rendering reveals that the transcriber could hear the Sanskrit cluster *-ddh-* and considered it worth

Table 9. *-p/t-s words ending in an obstruent coda in pre-ONWC/ONWC/STCA

Word	Pre-ONWC	ONWC (Corridor)	STCA	MTCA	CSZ/CollSZ
帶 OC *C.tˤat-s MC tajᴴ	—	*taC BZ: P. tac-/Skt. takṣ- Dk: tat-	*tɑi	*tɑi HL: 德奈, 當奈	CollSZ *tɑi T: da'i
害 OC *N-kˤat-s/ *m-kˤat-s MC hajᴴ	*ɣaC Dm: -[b] hās-	*ɣaC Dk: -has-	*ɣɑi	*ɣɑi HL: 何大, 何賴, 孩 蓋	CSZ: *hɑi~hɛi T, TD: he
賴 OC *rˤat-s MC lajᴴ	*laC Dm: -lat-	*laC BZ: P. -raṭ-/Skt. -rāṣṭ- Dk: rat-	*lɑi	*lɑi	CSZ: *lɑi~lɛi TD: le CollSZ *lɛi T: le'e
會 OC *m-kˤop-s MC hwajᴴ	—	*ɣuaC BZ: -vāsa	*ɣuɑi	*ɣuɑi HL: 胡外, 迴外	CSZ: *huɑi~huɛi C: hwa'ɨ
四 OC *s.li[t]-s (see Section 4.3) MC sijᴴ	—	*si BZ: -se-, P. -si, P. -s('e)-	*si, (*sit?) Jn: sid-	*si	CSZ: *si O, T: si; K, TD: sɨ; NT, DA: zi; Kbr: siysi, siysä, si

representing as -t + d- in the Chinese form (Coblin 1991: 70). By comparing the two renderings, one may infer that 四 contains -t in the underlying Chinese variety.

Words with *lˤ- and *-ʔ in Old Chinese

Of the four words with *lˤ- and *-ʔ mentioned in Section 3.4, three are employed in the transcriptional materials, namely 墮, 弟, and 道. As demonstrated in Table 10, they are not used to transcribe Indic th-. Other words with *lˤ- and *-ʔ that are found in the materials do not render Indic th- either, e.g. 殆 OC*lˤəʔ.

Interpreting the observations

The stages that are said to be associated with the city of Cháng'ān are the Cháng'ān dialect of ONWC, STCA, MTCA, and LTCA. LTCA data is virtually absent from the content above, and consequently the discussion below does not deal with this stage.

The Km data is silent on the words 天 "sky", 蛇 "snake", and 射 "shoot"; merely one example can be identified for the feature discussed in Section 3.4: 墮 "fall (v.)", which renders Indic d-, not th- (see Table 10). With respect to the words with *-p/t-s in Old Chinese, none of such words found in Kumārajīva's transcription contains a coda, e.g. 衛 "guard" OC *ɢʷrat-s > MC hjwejᴴ – ONWC *uei Km ve-. Judging from the limited data, the underlying language does not appear to be closely related to Old Western Chinese.

As STCA and MTCA refer to (two phases of) the Cháng'ān dialect spoken between Sui and Mid-Tang times, they are presumably identical to Old Western Chinese or at least a variety of it. For that matter, one would expect to be able to identify most, if not all,

Table 10. Selected words with *lˤ- and *-ʔ

Word	ONWC	STCA	MTCA	CSZ/CollSZ
墮 OC *lˤojʔ MC *dwa*^{X}	*duɑ BZ: -dvā-; Km: -dvā-; Dk: -dvā-	*duɑ	*duɑ HL: 徒果	CSZ *duɑ K, TD: dwa
弟 OC *lˤəjʔ MC *dej*^{X}	*dėi BZ: -tiṃ (=Gd. *-dhi-)	*dɨi	*dɨi > *diei Am: -dhe	CSZ *diei O: de
道 OC *kə.lˤuʔ MC *daw*^{X}	*dɑu	*dɑu	*dɑu HL: 陶老, 徒到	CSZ *dɑu K, T, TD, DA: de'u; TD: 'de'u CollSZ: *dɑu T: do'[u], da'o
殆 OC *lˤəʔ MC *doj*^{X}	*dɑi	*dɑi	*dɑi HL: 臺改	CSZ *dɑi~dɛi C: da'ɨ

Old Western Chinese features in the relevant materials. In reality, the Jn, Am, and HL data is not vastly different from Km data in this regard. Examples of the features in Section 3 are non-existent in Am/HL data. As for the transcription of Jñānagupta, 蛇 renders *yā* and 四 *sij*^{H} seemingly renders Indic *-d*. Nevertheless, concerning the latter case, it is crucial to note that other *-p/t-s words do not have a coda: ①寐 OC *mit-s > MC *mjij*^{H} – STCA *mɨ Jn: *-mi-*; ②致 OC *trit-s > MC *trij*^{H} – STCA *ti > ṭi Jn: *-ṭe ~ -ṭy(e)*, *-ṭy-*; ③肆 OC *s-ləp-s > MC *sij*^{H} – STCA *si Jn: *-si*. In light of this observation, it is unclear what to make of the suspicious example of 四. As a matter of fact, the near total absence of OC *-p/t-s > *-C is a bigger problem than the absence of any other feature for the STCA and MTCA periods on account of the fact that there are plenty of *-p/t-s words in the Jn/Am/HL data.

It then raises the possibility that the underlying languages of Jn, Am, and HL data were probably not the language of Guānzhōng at large. We speculate that the works may have been based on a variety to the east of Guānzhōng, which might have been the prestige variety then. Alternatively, it is worth considering that the eastern variety may have been spoken by a sizable portion of the residents of Cháng'ān, which can be regarded as Cháng'ān dialect per se.

With respect to the Gānsù Corridor, intriguingly, multiple features can be discerned in the Corridor dialect of ONWC. In the transcriptions of BZ/Dk, 天 renders Indic *h-* and 蛇 renders Indic *ya*. Moreover, words with *-aj*^{H} (< OC *-p/t-s) have an obstruent coda in both BZ and Dk, though unlike Old Western Chinese, *-p/t-s words having other Middle Chinese finals do not have such a coda: 致 OC *trit-s > MC *trij*^{H} – ONWC *ti Km: *-ti-*, *-ḍi*, *-ḍe*, *-ṭi*, *-li* Dk: *-ṭhe*. In spite of this difference, we suspect that the Corridor dialect is closely akin to Old Western Chinese; the preceding stage of Old Western Chinese (pre-OWC?) might have been introduced to the Corridor during or prior to the fourth century. Fast forward to the Shāzhōu dialects of the ninth and tenth centuries, no Old Western Chinese feature can be detected in them, as demonstrated in Tables 7–10. Here, it is pivotal to bear in mind that there is a large variety of materials for the Shāzhōu period, and thus the absence of the features cannot be mistaken. If one assumes that CSZ is the direct descendant of the Corridor dialect of ONWC, one convenient explanation to the discrepancy of the two stages is that the related forms, e.g. the *t*^{h}- variant of 天, were introduced from an easterly dialect. An alternative explanation would be that CSZ does not evolve from the Corridor dialect, and in my view, this is likely the case. First of all, the three features

together involve not only two to three words; for the third feature, ten -*aj*[H] words can be found in the CSZ/CollSZ materials, and none of the tokens contains a coda. Secondly, certain sound changes that occurred in mainstream Chinese in Medieval China are not observed in the Corridor dialect of ONWC but are reflected in CSZ, one of which is the labiodentalization of words with initials *Fēi*/*Fū*/*Fèng* 非/敷/奉母. If CSZ is the daughter language of the Corridor dialect of ONWC, then one is essentially positing that the change occurred individually in the Corridor; this scenario is not highly probable, as the conditions under which the sound change took place are rather specific from the perspective of the *Qièyùn* system: only words with a particular set of division-III finals are affected.

From a historical perspective, the Gānsù Corridor is a region which was, relatively speaking, rather sparsely populated. A major upheaval that caused significant numbers of people to flee the area could result in severe depopulation (Coblin 1994a: 13). And, conversely, even a modest influx of new settlers may well have had a momentous impact on the linguistic picture. Thus, it comes as no surprise that over the span of more than four centuries between the ONWC and CSZ periods, the original variety spoken in the region was supplanted by another variety introduced from outside the Corridor.

4.2 The time of divergence of OWC and Bai

The commonalities between Bai and Old Western Chinese reveal that they are closely related to each other, but it is unclear whether Bai is a sister language of Old Western Chinese or a daughter language of a particular stage of Old Western Chinese, e.g. the Tang stage. There exists some evidence to reject the departure of Bai during the Song dynasty. Beneath are extracts from three Song works (the first of which has appeared above):

(30) 《中山詩話》：……亦如關中以中為蒸，蟲為塵，丹青之青為萋也。
[*Zhōngshān shīhuà*: ... just as in Guānzhōng, 中 *trjuwng* is pronounced 蒸 *tsying*, 蟲 *drjuwng* is pronounced 塵 *drin*, and 青 *tsheng*, as in 丹青, is pronounced 萋 *tshej*.]

(31) 《老學庵筆記》卷六：四方之音有訛者，則一韻盡訛……秦人訛青字，則謂青為萋、謂經為稽。[*Lǎoxué ān bǐjì* Volume 6: Whenever there is inaccurate pronunciation in dialects, all words of the same rhyme are inaccurate.... The Qín people mispronounce 青, as such 青 is pronounced as 萋 and 經 (*keng*) as 稽 (*kej*).]

(32) 《耆舊續聞》卷七：關中人言清濁之清，不改「清」字；丹青之青，則為「萋」音。[*Qíjiù xùwén* Volume 7: For the word 清 *tshjeng*, people in Guānzhōng do not alter its pronunciation; but for 青, as in 丹青, it is pronounced 萋.]

These three extracts evince that words with -*eng* in Middle Chinese were pronounced -*ej* in the Guānzhōng dialect. In other words, the nasal coda was elided. This phenomenon is not observed in Bai (see Table 11).

The -*eng* words retain the nasal coda, with the exception of 暝 "[Bai] dark" *miɛ*[42], owing to the fact that *m*- and nasal vowels do not co-occur. It is no easy task to date the sound change in Old Western Chinese. Attestation of this feature in pre-Song western dialects is lacking, and thus it is tempting to suggest that this feature emerged in Song times. Nonetheless, ancient Chinese dialects were typically only scantily documented, especially the ones in earlier times, e.g. the Han dynasty; hence, the absence of attestation of it in earlier periods does not necessarily entail that the change had not occurred by the Tang dynasty. Regardless of when this sound change occurred, the fact that the elision of OC *-ŋ is not reflected in Bai at least shows that Bai is not a descendent of the Song stage of Old Western Chinese.

In a similar fashion, we know of no attestation of the features discussed in Sections 3.1–3.4 in the western dialects in Pre-Sui/Tang times (i.e. explicitly stated as a western

Table 11. Words with *-eng* in Middle Chinese

Word	OC	MC	Mandarin	Bai
青 "[Bai] blue"	*s.r̥ˤeŋ	*tsheng*	qīng	tɕʰɛ̃⁵⁵
星 "star"	*s-tsʰˤeŋ	*seng*	xīng	ɕɛ̃⁵⁵
瓶 "bottle"	*bˤeŋ	*beng*	píng	piɛ̃²¹
暝 "[Bai] dark"	–	*meng*[H]	míng	miɛ⁴²
聽 "hear"	*l̥ˤeŋ	*theng*	tīng	tɕʰɛ̃⁵⁵

feature), from the materials which survive today. As explicated above, one is incapable of concluding that they had not emerged by the Northern and Southern dynasties when dialectal information available to us is meagre in the first place. Until additional evidence emerges, dating the departure of Bai remains a challenging task.

4.3 The OC form of 四 "four"

The third issue is the reconstruction of the Old Chinese form of 四 "four" MC *sij*[H]. The pronunciations of 四 in Bai and Old Western Chinese have implications for the reconstruction of its Old Chinese form. Baxter and Sagart (2014) reconstruct the word as *s.li[j]-s, showing uncertainty between *-j-s and *-p/t-s, both Old Chinese sources of *-ij*[H]. In Zhengzhang's (2003) Old Chinese system, *-ij*[H] comes from *-(b/d/g)s or *-s (without a stop), tantamount to Baxter and Sagart's treatment; Zhengzhang rejected *-s and reconstructed 四 as *hljids. A number of pieces of evidence substantiate the rejection of *-j-s. 四 appears in one rhyme sequence in *Classic of Poetry*: Ode 53.1B (example (33)). In the sequence, the Old Chinese form of 紕 "furnish with an edge" is not available, but 畀 "give" has *-t-s in Old Chinese. 駟 "a vehicle drawn by four horses", the phonetic radical of which is 四, is also found in one rhyme sequence in *Classic of Poetry*: Ode 222.2B (example (34)); 㵑 "(flags) wave (v.)", 嘒 "chirp", and 屆 "arrive" are all reconstructed as *-t-s. These two rhyme sequences, especially the latter, lend support to *-p/t-s. Furthermore, that 四 ends in -t in Old Western Chinese and belongs to the EL words group in Bai implies that 四 likely has *-p/t-s in Old Chinese. As such, I put forward that 四 ends in *-p/t-s in Old Chinese. Following Baxter-Sagart's notation, the cluster can be denoted as *-[t]-s (i.e. *s.li[t]-s) when there is inadequate information to tell whether the cluster is *-p-s or *-t-s.

(33) 紕 "furnish with an edge" MC *bjie* > *pí*
      四 "four" OC *s.lij-s > MC *sij*[H] > *sì*
      畀 "give" OC *pit-s > MC *pjij*[H] > *bì*
(34) 㵑 "(flags) wave (v.)" OC *pʰˤit-s > MC *phej*[H] > *pì*
      嘒 "chirp" OC *qʷʰˤit-s > MC *xwej*[H] > *huì*
      駟 "a vehicle drawn by four horses" OC *s.lij-s > MC *sij*[H] > *sì*
      屆 "arrive" OC *kˤrit-s > MC *keaj*[H] > *jiè*

5. Conclusion

The present study explores a number of commonalities between Bai and Old Western Chinese. There exist words with *zy-* that are pronounced j- in both languages. The word for "fungus" in Old Western Chinese, 椹, is also found in Bai. Furthermore, the Old Chinese cluster *-p/t-s (> MOC *-t-s) yields -t in both Bai and Old Western Chinese.

Lastly, it appears that words with *lˤ- (whence MC *d*-) and *-ʔ in Old Chinese behave differently from other *d*- words in both languages. These observations, along with the commonalities spotted by other scholars, demonstrate that Bai and Old Western Chinese are closely related.

Abbreviations (excluding those introduced in Section 4.1)

ASP	Aspirated
EL	Entering-tone-like
MC	Middle Chinese
MOC	Middle Old Chinese
OC	Old Chinese
OWC	Old Western Chinese
SZ	Shāzhōu
UNASP	Unaspirated

References

Baxter, William H. 1992. *A Handbook of Old Chinese Phonology*. Berlin: Mouton de Gruyter.

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

Benedict, Paul K. 1982. "Sinitic and Proto-Chinese, part II: Bai and loans to Proto-Tai." Paper presented at the 15th International Conference on Sino-Tibetan Languages and Linguistics, Beijing, 17–19 August.

Brough, John. 1962. *The Gāndhārī Dharmapada*. London: Oxford University Press.

Chang, Kun. 1975. "Tonal developments among Chinese dialects", *Bulletin of the Institute of History and Philology* 46/4, 636–709.

Coblin, W. South. 1991. *Studies in Old Northwest Chinese*. Journal of Chinese Linguistics Monograph Series No. 4. Berkeley: University of California Press.

Coblin, W. South. 1992. "Comparative studies on some Tang-time dialects of Shazhou", *Monumenta Serica* 40, 269–361.

Coblin, W. South. 1994a. *A Compendium of Phonetics in Northwest Chinese*. Journal of Chinese Linguistics Monograph Series No. 7. Berkeley: University of California Press.

Coblin, W. South. 1994b. "Remarks on some early Buddhist transcriptional data from northwest China", *Monumenta Serica* 42, 151–69.

Downer, Gordon B. 1981. "Dialect information in the *Jiyun*." *Proceedings of the International Conference on Sinology: Section on Linguistics and Paleography*, 1–17. Taipei: Academia Sinica.

Gong, Xun. 2015. "How old is the Chinese in Bái? Reexamining Sino-Bái under the Baxter-Sagart reconstruction." Paper presented at the workshop "Recent Advances in Old Chinese Historical Phonology", London, 5–6 November.

Greenberg, Joseph H. 1953. "Historical linguistics and unwritten languages", in Alfred L. Kroeber (ed), *Anthropology Today*, 265–86. Chicago: University of Chicago Press.

Haudricourt, André-Georges. 1954a. "Comment reconstruire le chinois archaïque", *Word* 10/2–3, 351–64.

Haudricourt, André-Georges. 1954b. "De l'origine des tons en viêtnamien", *Journal Asiatique* 242, 69–82.

Lee, Man Hei. 2023. "Phonological features of Caijia that are notable from a diachronic perspective", *Journal of Historical Linguistics* 13/1, 82–114.

Lee, Yeon-Ju and Laurent Sagart. 2008. "No limits to borrowing: The case of Bai and Chinese", *Diachronica* 25/3, 357–85.

Matisoff, James 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, 24–28 October.

Pulleyblank, Edwin G. 1973. "Some further evidence regarding Old Chinese *-s* and its time of disappearance", *BSOAS* 36/2, 368–73.

Sagart, Laurent. 1981. "Aspiration-conditioned tone-lowering in Chinese dialects." Paper presented at the 14th International Conference on Sino-Tibetan Languages and Linguistics, Gainesville, Florida, 31 October – 1 November.

Starostin, Sergei. 1995. "The historical position of Bai", *МОСКОВСКИЙ ЛИНГВИСТИЧЕСКИЙ ЖУРНАЛ* [Moscow Journal of Linguistics] 1, 174–90.

Wang, Feng. 2006. *Comparison of Languages in Contact: The Distillation Method and the Case of Bai*. Language and Linguistics Monograph Series B: Frontiers in Linguistics III. Taipei: Institute of Linguistics, Academia Sinica.

Wang, Feng. 2013. *Hàn zàng yǔyán bǐjiào de fāngfǎ yǔ shíjiàn: hàn, bái, yí yǔ bǐjiào yánjiū* 漢藏語言比較的方法與實踐：漢、白、彝語比較研究 [Methods and implementations of Sino-Tibetan comparison: a comparative study of Chinese, Bai and Yi]. Beijing: Beijing daxue chubanshe.

Zhao, Jincan. 2010. "Yúnnán hèqìng báiyǔ yánjiū" 雲南鶴慶白語研究 [A study of Hèqìng Bai, Yúnnán], PhD dissertation, Minzu University of China.

Zhao, Yansun and Lin Xu. 1996. *Bái hàn cídiǎn* 白漢詞典 [Bai–Chinese dictionary]. Chengdu: Sichuan minzu chubanshe.

Zhengzhang, Shangfang. 1999. "Báiyǔ shì hànbái yǔzú de yīzhī dúlì yǔyán" 白語是漢白語族的一支獨立語言 [Bai is an independent branch of Sino-Bai], in Feng Shi and Wuyun Pan (eds), *Zhōngguó yǔyán xué de xīn tuòzhǎn: qìngzhù Wáng shìyuán jiàoshòu liùshíwǔ suì huádàn* 中國語言學的新拓展：慶祝王士元教授六十五歲華誕 [Recent advances in Chinese linguistics (celebrating Professor William S-Y. Wang's 65th birthday)], 19–73. Hong Kong: City University of Hong Kong Press.

Zhengzhang, Shangfang. 2003. *Shànggǔ Yīnxì* 上古音系 [Old Chinese phonology]. Shanghai: Shanghai jiaoyu chubanshe.

Zhengzhang, Shangfang. 2010. "Càijiā huà báiyǔ guānxì jí cígēn bǐjiào" 蔡家話白語關係及詞根比較 [Relationship between Caijia and Bai and comparison of their roots], in Wuyun Pan and Zhongwei Shen (eds), *Yánjiū zhī lè: qìngzhù Wáng shìyuán xiānshēng qīshíwǔ shòuchén xuéshù lùnwén jí* 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 75th birthday], 389–400. Shanghai: Shanghai jiaoyu chubanshe.

Zhengzhang, Shangfang. 2012. "Qín lǒng qùshēng wéi rù shízhèng" 秦隴去聲為入實證 [The evidence of the *qusheng* sounding like *rusheng* in Qin-Long dialect], *Yǔyán yánjiū* 語言研究 [Studies in language and linguistics] 32/3, 56–8.