17 On borrowing from Middle Chinese into Proto Tibetan: a new look at the problem of the relationship between Chinese and Tibetan[1]

MICHEL FERLUS

1 Introduction

It is well known that Chinese and Tibetan are genetically related. But it has been asserted by some scholars that an important part of the supposed common vocabulary represents an ancient layer of borrowings from Middle Chinese into Ancient Tibetan. It is important to separate the loanwords from inherited vocabulary in order to evaluate the closeness of the genetic relationship.

The study of borrowings depends on the relationship between the languages in contact. When the languages are not genetically related, borrowing usually concerns whole words, and it is in general relatively simple to sort out loanwords. But when genetically related languages are in contact, with some degree of intercomprehension, or at least a certain awareness of correspondences, and in a hierarchical relationship of prestige, more complex types of influence can occur. These may affect only one segment in a word, one constituent (consonant, vowel or rhyme) in a syllable, or one syllable in a dissyllabic word. Such phenomena have been observed by the author in situations of linguistic contact between Vietnamese and related languages within the Vietic linguistic group (Ferlus 1991, 1995,

---

[1] The following abbreviations are used:

MC Middle Chinese (Karlgren's Ancient Chinese): the stage of the $Qie^4 Yun^4$ reflected in the $Yun^4 Jing^4$.
OC Old Chinese (Karlgren's Archaic Chinese): the stage of the rhymes of the $Shi^1 Jing^1$
WT Written Tibetan.
PT Proto–Tibetan (The stage just before Middle Chinese influence).
PST Proto–Sino-Tibetan (in a restricted sense).
TB Tibeto-Burman.

2001). Meillet appealed to a phenomenon of this kind, under conditions of bilingualism, to explain the h in French haut (Germanic hoch, Latin altus) (Meillet 1936:99–103).

These phenomena will be illustrated here by borrowings, or partial borrowings, from Middle Chinese (MC) into Proto Tibetan (PT), the supposed stage of the language immediately preceding MC influence. The Proto Tibetan forms have been reconstructed by the author for the needs of the present article. This Proto Tibetan could just as well be called pre-Old Tibetan. The result of the Chinese influence is reflected in Old Tibetan and recorded in Written Tibetan (WT). It can be asserted that a part of PT vocabulary remained relatively close to Proto Sino-Tibetan (PST).

2   A theory of monosyllabisation from OC to MC

Before proceeding further, it is necessary to review the author's theory of the phenomenon of monosyllabisation that occurred between OC and MC (Ferlus 1998). This theory is used in the explanation of the influence of MC on Tibetan that follows.

Old Chinese was a disyllabic language, in the sense that while part of the vocabulary was monosyllabic, another part contained disyllabic words, more precisely of the sesquisyllabic type (as defined by Matisoff). This type is still widely represented in many Austroasiatic languages of Southeast Asia. A sesquisyllable is a type of disyllable composed of a main syllable preceded by a presyllable. The main syllable is similar to a monosyllabic word, while the presyllable is a reduced and unstressed syllable in which vocalic oppositions are neutralised. The presyllable can be a morphological prefix or a neutral element without any meaning.

monosyllable: CV(C)
sesquisyllable: C-CV(C)

According to my theory, OC sesquisyllables developed phonetic tenseness (T) while monosyllables developed laxness (L). Then, when sesquisyllabic words became monosyllabic by the loss of the presyllable, the earlier contrast of syllabic type, between C-CV(C) and CV(C), was replaced by the new contrast of tense vs lax (T/L). This phenomenon was associated with a vocalic split, with vowel lowering in T syllables and vowel raising in L syllables. Later, in a second step after these changes, the lenition of medial -r- further blurred the situation. This is the stage of MC characterised by the well-known system of four divisions: the T syllables belong to Division I/IV (syllables without medial -r- in OC) or to Division II (medial -r- in OC), while the L syllables belong to Division III (with or without medial -r- in OC), characterised by the famous yod of Karlgren's (1957) reconstructions.

Old Chinese (OC)		Middle Chinese (MC)	divisions		
C-CV(C)	*(tenseness)*	> CV(C) / T	*(vowel lowering)*	I/IV (–r) or II (+r)	
CV(C)	*(laxness)*	> CV(C) / L	*(vowel raising)*	III (± r)	

In my system, the symbol [ᵊ] is the mark of Division III. It indicates a raising and centralisation of the vowel associated with (what I suspect to be) breathy voice. The symbol ['] is the mark of Division II. It indicates the result of the lenition of OC medial -r-, probably a kind of velar spirant. Thus Division II is simply an offshoot of Division I (and IV, below); together these three divisions continue the old T category. In L syllables, the softened OC medial -r- became obscured by the breathiness of the vowel and was lost; there was no split

analagous to the one that gave rise to Division II. Thus the entire L category is continued intact by Division III. No special mark characterises Division I or IV. Division IV is in complementary distribution with Division I, apparently a device to represent the single MC front diphthong. This theory is summarised in Table 1.

Table 1: Proposed origins of MC divisions

syllable-type	OC medial (Baxter)	MC division
(without medial -r-)		
tense	Ø	**I/IV** / T
lax	-j-	**III** / L
(with medial -r-)		
tense	-r-	**II** / T(r)
lax	-rj-	**III** / L(r)

The examples in Table 2 are taken from Baxter (1992). I have added my own phonetic interpretation between square brackets.

Table 2: Diachronic examples of syllable-types with phonetic interpretation

T/L	div.		Man.	MC	OC
T	I	納	na^4	nop [nʌp]	*nup [^{T}(C-)nup] 'send in' (695h)
L	III	入	ru^4	nyip [n̪ᵊip]	*n-j-up [^{L}nup) 'enter' (695a)
T	IV	銘	$ming^2$	meng [miɛŋ]	*meng [^{T}(C-)meŋ] 'inscription' (826d)
L	III(>IV)	戌	$ming^2$	mjieng [mᵊjeŋ]	*m-j-eng [^{L}meŋ] 'name' (826a)
T	I	股	gu^3	kuX [kɔʔ]	*kaʔ [^{T}(C-)kaʔ] 'thigh' (51a)
T(r)	II	假	jia^3	kæX [kᴶæʔ]	*k-r-aʔ [^{T}(C-)kraʔ] 'false, simulate' (33c)
L(r)	III	莒	ju^3	kjoX [kᵊʌʔ]	*k-rj-aʔ [^{L}kraʔ] 'round basket' (76j)

Before the complete, structural monosyllabisation that affected the whole sesquisyllabic vocabulary, there may have existed a slower process of random monosyllabisation affecting individual words.

3   The mode of borrowing from Middle Chinese into Proto Tibetan

Language A (here MC) is in a dominating position with a genetically related language B (here PT). Language A is regarded as prestigious by speakers of B who, by a kind of affectation, are led to imitate some characteristic sounds of A unknown in B. This results in a phonetic compromise, a segment of an A word being borrowed and substituted for the corresponding segment of the cognate B word. The sounds of MC that did not exist in PT are the segments (rhymes or main syllables) that characterise Division III (i.e. the presumed breathiness marked by [$]) and Division II (i.e. the spirantised velar sound marked by [ɦ]).

Speakers of PT tried, unconsciously or not, to imitate these unfamiliar sounds which were felt to carry with them the prestige of the dominating language. But in contrast to the usual process, in which whole words are borrowed, only the the characteristic MC segments of cognate words were borrowed by speakers of PT.

Table 3: Chinese and Tibeto-Burman numerals

		Karlgren 1957	Coblin 1986	Coblin 1986	Coblin 1986	Benedict 1972	Pulleyblank 1991
		Ar C > An C	PST	OC > MC	TB	TB	EMC
1	隻 zi^1	--- > tśi̯äk	gtyik	tjik > tśjäk	g-tyik	t(y)ik	tɕiajk
2	二 er^4	ńi̯ər > ńźi-	gnyis	njidh > ńźi	gnyis	g-nis	ɲiʰ
3	三 san^1	səm > sâm	gsum	səm > sâm	g-sum	g-sum	sam
4	四 si^4	si̯əd > si-	bɫyid	sjidh > si-	blyiy	b-liy	siʰ
5	五 wu^3	ngo > nguo:	lngaɣ	ngagx > ngwo:	l-nga	l-ŋa	ŋɔʔ
6	六 liu^4	li̯ôk > li̯uk	dljəkw	ljəkw > ljuk	d-ruk	d-ruk	luwk
7	七 qi^1	ts'i̯ĕt > ts'i̯ĕt	shnjis	tshjit > tshjet	s-nis	s-nis	tsʰit
8	八 ba^1	pwăt > pwăt	priat	priat > pwăt	pryat	b-r-gyat	pəɨt/pɛːt
9	九 jiu^3	ki̯ŭg > ki̯ə̯u:	dkwjəɣw	kjəgwx > kjəu:	d-kuw	d-kuw	kuwʔ
10	十 shi^2	d̑i̯əp > źi̯əp	gřip	djəp > źjəp	gip	gip	dʑip
100	百 bai^3	păk > pɒk	pria?	prak > pɐk	prya	r-gya	paɨjk/pɛːjk

Table 3: (continued)

Baxter 1992	Ferlus 1998	Sagart 1999	Matisoff 1997	WT	
OC > MC	OC > MC	OC > MC	PST		
(tjek > tsyek)	ᴸtek > tɕˀek	-------	g-t(y)i-k ⋊ tya-k	*gcig*	1
(njits > nyijH)	ᴸnits > n̪ˀijʰ	ᵇni[jt]-s > nyijH	g-ni-s/k	*gnyis*	2
sum > [sam]	ᵀksum > [sam]	ᵃs-hlɨm > sam	g-sum	*gsum*	3
s(p)jij/ts > sijH	ᴸslits > sˀijʰ	ᵇs-hli[j]-s > sijH	b-liy = b-ləy	*bzhi*	4
ngaʔ > nguX	ᵀlŋaʔ > ŋɔʔ	ᵃŋaʔ > nguX	l-ŋa ⋊ b-ŋa	*lnga*	5
C-rjuk > ljuwk	ᴸruk > lˀuwk	ᴮCə-ruk > ljuwk	d-ruk / d-k-rok	*drug*	6
thsjit > tshit	ᴸtsʰit > tsʰˀit	ᵇs-hnit > tshit	s-nis	(*bdun*)	7
(pret > pɛt)	ᵀpret > pᴶɛt	ᵃpr[e]t > peat	b-r-gyat ⋊ b-g-ryat	*brgyad*	8
kʷjuʔ > kjuwX	ᴸkʷuʔ > kˀuwʔ	ᵇkuʔ > kjuwX	d-kəw ⋊ s-gəw ⋊ d-gaw	*dgu*	9
gjip > dzyip	ᴸgip > dʑˀip	ᵇgip > dzyip	gip ⋊ gyap	*bcu*	10
prak > pæk	ᵀprak > pᴶæk	--------	b-r-gya ⋊ b-g-rya	*brgya*	100

This process will be illustrated first in the numerals. Table 3 shows the principal reconstructions and interpretations of the set of numerals 'one' to 'ten' and 'hundred': Archaic Chinese and Ancient Chinese (Karlgren 1957), their equivalents Old Chinese and

Middle Chinese (Coblin 1986, Baxter 1992, Ferlus 1998, Sagart 1999), Early Middle Chinese (Pulleyblank 1991), Proto Sino-Tibetan (Coblin 1986, Matisoff 1997), and Tibeto-Burman (Benedict 1972, Coblin 1986).

4 Comparision of Tibetan and Chinese numerals

Table 4 is the reference chart for the following discussion. The reconstructions used here are by the author. For OC and MC they are based on Baxter 1992. For PT and PST they have been elaborated for the needs of the present article.

Table 4: Tibetan numerals

			OC > MC	PT + MC	*hypothetic*	WT
1	隻	zhi^1	^{L}tek > tɕ^{ə}ek	k[tek] + tɕ^{ə}ek	> ktɕ^{ə}ek	*gcig*
2	二	er^4	^{L}nits > ɳ^{ə}ij^h	k[nits] + ɳ^{ə}ij^h	> kɳ^{ə}ij^h	*gnyis*
3	三	san^1	^{T}ksum > [sam]	ksum	> (*unchanged*)	*gsum*
4	四	si^4	^{L}slits > s^{ə}ij^h	p[sits] + s^{ə}ij^h	> ps^{ə}ij^h	*bzhi*
5	五	wu^3	^{T}lŋaʔ > ŋɔʔ	lŋa	> (*unchanged*)	*lnga*
6	六	liu^4	^{L}ruk > l^{ə}uwk	truk	> (*unchanged*)	*drug*
7	七	qi^1	^{L}tsh^{ə}it > tsh^{ə}it			(*bdun*)
8	八	ba^1	^{T}pret > p^{J}ɛt	pr[et + p]^{J}ɛt	> pr^{J}ɛt	*brgyad*
9	九	jiu^3	^{L}k^{w}uʔ > k^{ə}uwʔ	t[ku] + k^{ə}uwʔ	> tk^{ə}uwʔ	*dgu*
				or tku	> (*unchanged*)	
10	十	shi^2	^{L}gip > dz^{ə}ip	p[gip] + dz^{ə}ip	> dz^{ə}i(p)	*bcu*
100	百	bai^3	^{T}prak > p^{J}æk	pr[ak + p]^{J}æk	> pr^{J}æ(k)	*brgya*

'One ~ alone':
  PT 'one' *ktek, WT *gcig*.
  OC 'alone' (tjek) [^{L}tek] > MC (tsyek) [tɕ^{ə}ek] > zhi^1 隻 (1260c), not cited in
      Baxter (1992). The current word for 'one' is yi^1 < MC ʔjit [ʔ^{ə}it] < OC *ʔjit [^{L}ʔit].
  PST *ktek.

The presyllable k-, reconstructed on the basis of WT *g*-, was lost in pre-OC times by random monosyllabisation. The division III of MC requires the reconstruction of a monosyllable in OC.

As part of the interference of MC forms with PT forms, the main syllable in PT *ktek was replaced by the unfamiliar pronounciation for Tibetan speakers of MC [tɕ^{ə}ek]. The combination *k[tek] + tɕ^{ə}ek gave rise to the hypothesised intermediate form *ktɕ^{ə}ek, well represented by WT *gcig*.

'Two':
  PT *knits, WT *gnyis*.
  OC (*njits) [^{L}nits] > MC (nyij^H) [ɳ^{ə}ij^h] > er^4 二 (564a), not cited in Baxter.
  PST *knits.

The presyllable k-, reconstructed on the basis of WT *g*-, was lost in pre-OC times by random monosyllabisation. MC division III points to an OC monosyllable. The final -ts changed into -js > -jʰ by final cluster simplification (Baxter 1992:568–9).

The main syllabe in PT *knits was replaced by the unfamiliar pronounciation of MC n̠ˀijʰ. The combination *k[nits] + n̠ˀijʰ (with the possibility of a pre-MC form n̠ˀis) gave rise to the hypothesised intermediate form *kn̠ˀijʰ, well represented by WT *gnyis*.

'Three':
  PT *ksum, WT *gsum*.
  OC *sum [ᵀksum] > MC sam [sam] (irregular rhyme) > *san¹* 三 (648a). The regular MC rhyme is [-ʌm] (Baxter -om).
  PST *ksum.

The pronunciations of MC sam as well as any other MC forms in -ʌm (see below), all belonging to the divisions I or II, were not exotic for Tibetan speakers. So no imitation occurred, and WT *gsum* derives directly from PT *ksum without interference from MC.

The problem raised by the reconstruction of 'three' and its word family is a very complex one. It has been treated in detail by Sagart (1999:14F152). A detailed discussion here would lead us too far from the present subject. I will only briefly outline my point of view.

Sagart proposed two forms for 'three': OC *ᵃs-hlɨm > MC sam > *san¹* 三 for the simple graph and OC *ᵃs-hlɨm > MC tshom > *can¹* 参 for the complex graph (or *da⁴ xie³*). First of all, I consider that the rhymes -um / -up must be reconstructed, and that the changes -um > -ɨm and -up > -ɨp occured after OC times. Aside from the basic form OC ᵀksum > *san¹* 三 'three', the word family compris es the MC meaning o 骖 (read *can¹*), MC tshom [tshʌm] 'three horses in a team', and both MC meanings of 参 (read *can¹*): MC tsho [tshʌm] 'three, a triad' and (read *shen¹*) MC srim [ʂˀim] 'the triad star of Orion'. The character 参 (read *san¹*) is also used even today as a complex graph for 'three'. It must be noted that MC rhymes in tshom [tshʌm] (division I) and in srim [ʂˀim] (division III) are regular in respect to the OC rhyme -um [-um]. For these two words I propose the reconstructions OC *srum [ᵀksrum] > MC tshom [tshʌm] and OC *srjum [ᴸsrum] > MC srim [ʂˀim], which I consider as secondary forms of OC ᵀksum. To summarise (with my reconstructions only):

  pre-OC ksum > OC ᵀksum > MC sam (irr.) > *san¹* 三 'three'.
  pre-OC krsum > (metathesis of -r-) OC ᵀksrum > MC tshʌm 参 'three, a triad' > *can¹*
      (also, incidentally, the reading with the meaning 'take part, visit'), also MC tshʌm 骖
      'three horses in a team'.
  pre-OC krsum > (loss of k- and metathesis of -r-) OC ᴸsrum 参 'the triad star of
      Orion' > MC ʂˀim > *shen¹* (also the reading with the meaning 'ginseng').

The change of pre-OC krsum into OC ᵀksrum or ᴸsrum by metathesis of -r- from the presyllable to the main syllable is, of course, purely hypothetical. But the phenomenon of metathesis, although refused by some scholars (Handel 2002), can help us to understand the curious instability and the intrusive behaviour of some OC medial -r-, and the fact that items with or without this medial can occur in the same phonetic series. The idea of an ancient metathesis of -r- is supported by some lexical correspondences between WT and OC. Compare WT *rdul* 'dust' with *chen2* 塵 < MC drin [ɖˀin] < OC *drjɨn [ᴸdrɨn] (< pre-OC *drɨl) 'id.' (example from Coblin 1986:68).

I propose that the archaic character for 參 previously had the meaning 'three horses in a team' because this notion was more familiar than 'triad of Orion'. The meaning of three horses is expressed by the upper part of the archaic character, which rather clearly shows the three horse's heads, contra some other scholars who prefer to see three stars in it (why should the three stars be tied on?). The lower part of the character has sometimes been interpreted as the phonetic element, but neither the element *zhen³* 抮 (OC rhyme -ɨn) nor *shan¹* 杉 (OC rhyme -am) fits phonetically with *can¹* 參 (OC rhyme -um). For myself, I prefer to see in the lower part of the archaic character for 參 the image of reins hung with ornaments.

The ancient pronunciation of 參 was used to derive numerous other characters that belong to the phonetic series GSR 647.

'Four':
    PT *psits, WT *bzhi*.
    OC *s(p)jij/ts [^{L}slits] > MC sij^H (s^{ə}ij^h) > *si⁴* 四 (518a).
    PST *plsits ~ *pslits.

The presyllable p- (perhaps a prefix ?), reconstructed on the basis of WT *b-*, was lost in pre-OC times. The medial -l- is justified by occurrences in some Tibeto-Burman languages.

The main syllable in PT psits was replaced by a corrupted form of the unfamiliar pronunciation of MC s^{ə}ij^h. The combination *p[sits] + s^{ə}ij^h gave rise to the hypothesised intermediate form *ps^{ə}ij^h, which is rather well represented by WT *bzhi*.

'Five':
    PT *lŋa, WT *lnga*.
    OC *ngaʔ [^{T}lŋaʔ] > MC ngu^X [ŋɔʔ] > *wu³* 五 (58.
    PST *lŋaʔ.
        Like 'three', WT *lnga* derives directly from PT.

'Six':
    PT *truk, WT *drug*.
    OC *C-rjuk [^{L}ruk] > MC ljuwk [l^{ə}uwk] > *liu⁴* 六 (1032a).
    PST *truk ~ t-ruk.

The first element t- was lost during pre-OC times, suggesting that tr- must have been a kind of disjoined cluster. In Proto Thai proper, 'six' is reconstructed as *^{h}rok, the voicelessness being irrefutable proof of the presence of an old presyllabic element. As in the case of 'three' and 'five', the WT form derives directly from PT.

'Seven':
    PT *pdun, WT *bdun*.
    OC *thsjit [^{L}tsʰit] > MC tshit [tsʰ^{ə}it] > *qi¹* 七 (400a).
    Pre-OC (for PST) *snit ~ sʰnit. Note that the rhyme -it(s) exits in knits 'two',
        plsits ~ pslits 'four (2+2)' and snit ~ sʰnit 'seven (5+2)'.
        The Tibetan and Chinese forms are not genetically related.

'Eight':
    PT *pret, WT *brgyad*.
    OC (*pret) [^{T}pret] > MC (pɛt) [p^{ɹ}ɛt] > *ba¹* 八 (281a).
    PST *pret.

The rhyme -et in PT pret was replaced by the unfamiliar pronounciation of the segment -ᴵɛt of MC pᴵɛt. The combination *pr[et + p]ᴵɛt gave rise to the hypothesised intermediate form *prᴵɛt, rather well represented by WT *brgyad*. The segment -*gyad* is the result of the interpretation of -ᴵɛt in the phonetic system of Tibetan. Some scholars have interpreted -*g*- as an epenthetic element, but it must be remarked that epenthesis normally occurs at the junction of two syllables; it did not occur in the Tibetan word for 'eight'.

'Nine':
   PT *tku, WT *dgu*.
   OC *kʷjuʔ [ᴸkʷuʔ ~ ᴸkuʔ] > MC kjuwˣ [kᵊuwʔ] > *jiu³* 九 (992a). Baxter's reconstruction of a labiovelar before a high rounded vowel is surprising, in spite of his solid argument, so I will propose an alternate form ᴸkuʔ.
   PST *tkuʔ ~ *tku.

At first sight, WT *dgu* appears to derive directly from PT *tku, but the voiced velar -*g*- does not fit perfectly with PST and could be a result of the laxness of the MC form. If so, the combination *t[ku] + kᵊuwʔ gave rise to the hypothesised intermediate form *tkᵊuwʔ represented by WT *dgu* .

'Ten':
   PT *pgip (?), WT *bcu*.
   OC *gjip [ᴸgip] > MC dzyip [dʑᵊip] > *shi²* 十 (686a)
   PST *pgip.

The presyllable p- (perhaps a prefix ?) is reconstructed on the basis of WT *b*- assuming that these forms are related. I suppose that the combination *p[gip] + dʑᵊip gave rise to the hypothesised intermediate form *pdʑᵊip. Could the WT rhyme -*cu* represent MC dʑᵊip after the loss of final -p? I must confess that I am not sure.

'Hundred':
   PT *prak, WT *brgya*.
   OC *prak [ᵀprak] > MC pæk [pᴵæk] > *bai³* 百 (781a).
   PST *prak.

The demonstration for 'hundred' is parallel to that for 'eight'. The combination *pr[ak + p]ᴵæk gave rise to the hypothesised intermediate form *prᴵæk > prᴵæ, rather well represented by WT *brgya*. The loss of final -k is unexplained but not unprecedented.

Among the eleven comparisons between Tibetan and Chinese numerals, nine can be considered as good correspondences. The words for 'seven' are not cognate, and the correspondence for 'ten' is not absolutely sure. In the correspondences for 'three', 'five' and 'six', the WT forms derive directly from PT without MC interference. They must be considered as pure inherited correspondences. In the correspondences for 'one', 'two', 'four', 'nine' (possibly) and 'ten' (if related), the MC monosyllable replaced the main syllable in the PT form; in the case of 'eight' and 'hundred' the segment replaced was the rhyme. These are what could be called corrupted (or modified) inherited correspondences, in which the Tibetan word is the result of a compromise between an inherited form and a borrowed segment while the Chinese term remains unchanged. These modified correspondences fall between pure inherited correspondences and full borrowing.

5 Additional examples

In this section, further examples are presented to illustrate both corrupted and regular correspondences between Tibetan and Chinese. They aren taken from Coblin (1986) and Gong (1995). OC and MC forms are between square brackets; PT and PST forms are the author's. OC and MC forms follow the system of Baxter (1992); when not attested they are placed between brackets. We begin with examples where the inherited correspondence has been perturbed by direct influence.

'Weary, exhausted':
PT *bral, WT 'fatigue, weariness' *o-brgyal*, 'to faint' *brgyal*.
OC (*brjaj) [bral > ^{L}braj] > MC (bje) [bᵊe] > *pi²* 疲 (25d), also *pi²* 罷 (26a).
PST *bral.

It seems here that the influence extends only to the vowel of the PT form: *br[a]l + [b]ᵊe gave rise to the hypothesised intermediate form *brᵊal, represented by WT *brgyal*.

'Dwell, etablish':
PT *bdoks, WT 'to sit, dwell' *bzhugs-pa.*
OC (*djoʔ) [^{L}doʔ] > MC dzyu^{X} [dzᵊuʔ] > *shu⁴* 樹 (127j).
PST *bdoks.

It seems that the segment -do- of PT *bdoks was corrupted by the MC form according to the formula *b[do]ks + dzᵊuʰ, the result being represented by WT *bzhugs*.

'To flow, flowing':
PT *run, WT *rgyun.*
OC (*wrjɨn) [^{L}wrɨn] > MC (hwin) [ɣwᵊin] > *yun²* 沄 (see GSR 460 and 227).
  The character 沄 given by Coblin (1986) is not attested in Karlgren (1957), Pulleyblank (1991), or Baxter (1992).
PST *(C)rɨ/un (vowel reconstruction uncertain).

The rime of PT *run was influenced by the rime of MC: r[un] + [ɣw]ᵊin, obviously at a stage earlier than MC proper, before the OC medial -r- merged into the breathiness of the division III lax syllable. See 'eight', OC (*pret) [^{T}pret] > MC (pet) [pᴶɛt] > *ba¹* 八, WT *brgyad*. In any case the segment *-gyun* of WT cannot be an inherited form. This correspondance shows what Bodman (1980) called 'primary yod', considered today as indicating an acquired correspondence.

'Center, middle':
PT *gruŋ, WT 'middle, midst' *gzhung.*
OC (*k-ljung) [^{L}truŋ] > MC (trjuwng) [ʈᵊuwŋ] > *zhong¹* 中 (1007a).
PST *kruŋ/truŋ.

The segment -ruŋ of PT *gruŋ was replaced by MC ʈᵊuwŋ according to the formula *g[ruŋ] + ʈᵊuwŋ with an approximative phonetic adjustment, the result being represented by WT *gzhuang*.

'Loyal, sincere':
PT *gruŋ, WT 'to attend to, sincere' *gzhung.*
OC *[^{L}truŋ] > ^{L}luŋ] > MC [ʈᵊuwŋ] > *zhong¹* 忠 (1007k).
  The OC and MC forms are set up on the model of *zhong¹* §§ (1007a).
PST *kruŋ/truŋ.

The process is the same as for 'center, middle'.

'Salt, salty':
PT *ram, WT *rgyam-tshwa.*
OC (*(C)-rjam) [(C)ram > ^{L}ram] > MC (yem) [j^{ə}em] > *yan^2* 鹽 (609n).
   The MC initial is irregular; according to the basic phonetic -ram of GSR 609, the
   regular initial would be l-.
PST *(C)ram.

The rime of PT ram was replaced by an earlier form of MC: r[am] + j^{ə}em, the result being reinterpreted by WT *rgyam*. For the interpretetion of OC -r- by WT *-gy-* see 'eight' and 'hundred' above, 'weary, exhausted' and 'to blow'.

'To see':
PT *mk^{h}en, WT *mkhyen-pa.*
OC *kens [^{T}kens] > MC ken^H [kiεn^h] > *jian^4* 見 (241a). The MC vowel -iε- is
   characteristic of division IV.
PST *(C)ken. A presyllable must be reconstructed to explain divisions I/IV.

The rime of PT *mk^{h}en was replaced by the rime -iεn of the MC form according to the formula *mk^h[en] + [k]iεn[s], the result being represented by WT *mkhyen*.

'Imitate, conform to':
PT *sbaŋ, WT 'to learn, study, exercise' *sbyong(s), sbyang(s).*
OC 'imitate' *pjang? [^{L}paŋ?] > MC pjang^X [p^{ə}aŋ$^?$] > *fang^3* 放 (740i); also OC
   'method, norm' *pjang [^{L}paŋ] > MC pjang [p^{ə}aŋ] > *fang^1* 𣃐 (740a).
PST *spaŋ.

The rime of PT *sbaŋ was replaced by the rime of the MC form: *sb[aŋ] + [p]^{ə}aŋ, the result being represented by WT *sbyong/sbyang*.

'Taste':
PT *snep (?), WT *snyab-pa.*
OC (*snep > *thep) [snep > ^{T}t^{h}ep] > MC (thep) [t^{h}iεp] > *tie^1* 呫 (618p). The MC
   vowel -iε- is characteristic of division IV.
PST *snep.

Coblin (1986) proposed OC hniap > MC thiep and reconstructed PST sniap, in fact on the basis of the WT form.

The rime of the supposed PT snep was replaced by the rime -iεp of the MC form according to the formula *sn[ep] + [t^h]iεp, with palatalisation of the nasal, the result being represented by WT *snyab*. This case is similar to that of *mkhyen* 'to see'.

By way of contrast, we present below some correspondences involving OC medial -r- in which the two languages have evolved independently without interference (as in the cases of 'three', 'five' and 'six'). These are considered to be inherited correspondences, to be compared with the corrupted correspondences above.

'Add, apply':
PT *pkral, WT 'to impose, to appoint to' *bkral.*
OC (*kral >) *kraj [kral > ^{T}kraj] > MC kæ [k^{J}æ] > *jia^1* 加 (15a).
PST *pkral.

'Bear, rear':
  PT *srel, WT 'bring up, rear' *srel.*
  OC (*srel >) *srjen [*srel > ^Tsren] > MC srɛn [ş^Jɛn] > *chan³* 產 (194a).
  PST *srel

'Busy, employed at':
  PT *brel, WT *brel.*
  OC (*brels) [^Tbrels] > MC (bɛn^H) [b^Jɛn^h] > *ban⁴* 辦 (219f).
  PST brel.

'Hard, strong':
  PT *kraŋ, WT 'hard' *khrang.*
  OC 'strong' *kraŋʔ [^Tkraŋʔ] > kængˣ [k^Jæŋʔ] > *geng³* 梗 (745e).
  PST *kraŋ

'Shell, armor':
  PT *krap, WT 'shield, coat of mail' *khrab.*
  OC *krap [^Tkrap] > MC kæp [k^Jæp] > *jia³* 甲 (629a).
  PST *krap

'Weep':
  PT *krap, WT 'a weeper' *khrab-khrab.*
  OC *krjɨp [^Lkrɨp] > MC khip [k^əip] > *qi⁴* 泣 (694h).
  PST *krap

6   Conclusion

We observe that the Tibetan language shows a two widely differing types of phonetic developments from PST and PT: one is straightforward and can be considered as regular while the other can be regarded as abnormal.

The regular changes, where the rimes are well preserved, can be represented by examples such as 'three' (*ksum > *gsum*), 'add, apply' (*pkral > *bkral*) and 'weep' (*krap > *khrab*), showing clear correspondences with OC. Other examples are: 'five' (*lŋa > *lnga*), 'six' (*truk > *drug*), 'nine' (*tku > *dgu*), 'bear' (*srel > *srel*), 'busy' (*brel > *brel*), 'hard' (*kraŋ > *khrang*) and 'armor' (*krap > *khrab*). However, some of these may be pure borrowings of the whole words from MC into PT.

The abnormal changes, in which the rimes have been corrupted, can be represented by such examples as 'eight' (*pret > *brgyad*), 'weary, exhausted' (*bral > *brgyal*) and 'salt' (*ram > *rgyam*), showing irregular correspondences with OC. Other examples are: 'one' (*ktek > *gcig*), 'two' (*knits > *gnyis*), 'four' (*psits > *bzhi*), 'dwell' (*bdoks > *bzhugs*), 'to flow' (*run > *rgyun*), 'center' (*gruŋ > *gzhung*), 'loyal' (*gruŋ > *gzhung*), 'salt' (*ram > *rgyam*), 'to see' (*mkʰen > *mkhyen*), 'imitate' (*sbaŋ > *sbyong*) and 'taste' (*snep > *snyab*). It is precisely to explain such correspondences that the author has proposed the borrowing of segments from Middle Chinese forms into Proto Tibetan as detailed above.

This mode of borrowing proposed, in which only a part of the word (main syllable or rhyme) is affected, is the consequence of a particular situation. The two languages in contact are genetically related, with a certain degree of intercomprehension and in a hierarchical relation of prestige. The Chinese language of MC times, being in a dominating position,

was regarded as prestigious by speakers of Tibetan, who were led to imitate, by a kind of affectation, the characteristic features of Divisions II and III which were unknown in Tibetan.

Consideration has been limited here to a restricted domain, mostly the numerals, which behave in general as a group, but the analysis could be extended with profit to most of the Tibetan vocabulary. I propose to call this special process of borrowing 'hypercorrection by affected imitation'. Such layers of borrowing have never been clearly identified in historical comparative studies.

References

Baxter, William H., 1992, *A handbook of Old Chinese phonology.* Berlin, New York: Mouton de Gruyter.

1993, Some phonological correspondences between Chinese and Tibeto-Burman. Paper presented to the 26th International Conference on Sino-Tibetan Languages and Linguistics, Osaka, Japan.

Benedict, Paul K., 1972, *Sino-Tibetan: a conspectus.* Contributing editor, James A. Matisoff. Cambridge: Cambridge University Press.

1976, Sino-Tibetan: another look. *Journal of the American Oriental Society* 96:167–197.

Bodman, N.C., 1980, Proto-Chinese and Sino-Tibetan: data towards establishing the nature of the relationship. In Franz van Coetsem and Linda Waugh eds *Contributions to historical linguistics: issues and materials*, 34–199. Leiden: Brill.

Coblin, W. South, 1986, *A Sinologist's handlist of Sino-Tibetan lexical comparisions.* Monumenta Serica Monograph Series XVIII. Nettetal: Steyler.

Ferlus, Michel, 1991, Le dialecte vietnamien de Vinh. Paper presented at 24th International Conference on Sino-Tibetan Languages and Linguistics, Ramkhamhaeng University, Bangkok, Thailand.

1995, Particularités du dialecte vietnamien de Cao La Ha (Quang Binh, Vietnam). Paper presented at dixièmes journées de linguistique de l'Asie Orientale, Ecole des Hautes Etudes en Sciences Sociales, Paris.

1998, Du chinois archaïque au chinois ancien: monosyllabisation et formation des syllabes *tendu / lâche* (Nouvelle théorie sur la phonétique historique du chinois). Paper presented at 31st International Conference on Sino-Tibetan Languages and Linguistics, University of Lund, Sweden.

2001, Les hypercorrections dans le thô de Lang Lo (Nghê An, Vietnam): les pièges du comparatisme. Paper presented at quinzièmes journées de linguistique de l'Asie Orientale, Ecole des Hautes Etudes en Sciences Sociales, Paris.

Gong Hwang-cherng, 1980, A comparative study of the Chinese, Tibetan and Burmese vowel system. *Bulletin of the Institute of History and Philology*, Academia Sinica 51/3:455–490.

1995, The system of finals in Proto-Sino-Tibetan. In William S-Y. Wang ed. *The ancestry of the Chinese language*, Journal of Chinese Linguistics Monograph Series No. 8, 41–92.

Handel, Zev., 2002, Rethinking the medials of Old Chinese: where are the *r*'s? *Cahiers de Linguistique Asie Orientale* 31/1:3–32.

Karlgren, Bernhard, 1957, Grammata Serica recensa. Reprint from *Bulletin of the Museum of Far Eastern Antiquities* 29:1–332.

Matisoff, James A., 1997, *Sino-Tibetan numeral systems: prefixes, protoforms and problems*. Canberra: Pacific Linguistics.

Meillet, Antoine, 1936, Sur le bilinguisme. *Linguistique historique et linguistique générale*, vol. II. Collection Linguistique, Société Linguistique de Paris, 40:99–103. Paris: Klincksieck.

Miller, Roy Andrew, 1988, The Sino-Tibetan hypothesis. *Bulletin of the Institute of History and Philology*, Academia Sinica 59/2:509–540.

Norman, Jerry and W. South Coblin, 1995, A new approch to Chinese historical linguistics. *Journal of the American Oriental Society* 115/4:576–584.

Peiros, Ilia and Sergei A. Starostin, 1996, *Comparative dictionary of five Sino-Tibetan languages*. Melbourne: University of Melbourne Department of Linguistics and Applied Linguistics.

Pulleyblank, Edwin G., 1973, Some new hypotheses concerning word families in Chinese. *Journal of Chinese Linguistics* 1/1:111–125.

1991, *Lexicon of reconstructed pronunciation in Early Middle Chinese, Late Middle Chinese and Early Mandarin*. Vancouver: University of British Columbia.

1998, Old Chinese and Sino-Tibetan. Paper presented at 31st International Conference on Sino-Tibetan Languages and linguistics, University of Lund, Sweden.

Sagart, Laurent, 1995a, Some Remarks on the Ancestry of Chinese. In William S-Y. Yang ed. *The ancestry of the Chinese language*. Journal of Chinese Linguistics Monograph Series No. 8, 195–223.

1995b, Questions of method in Chinese-Tibeto-Burman comparision. *Cahiers de Linguistique Asie Orientale* 24/2:245–255.

1999, *The roots of Old Chinese*. Amsterdam, Philadelphia: John Benjamins.

Schuessler, Axel, 1974, *R* and *L* in Archaic Chinese. *Journal of Chinese Linguistics* 2/2:186–199.

Starostin, Sergei A., 1996, Vowel correspondances in Sino-Tibetan. Paper presented to the 29th International Conference on Sino-Tibetan Languages and Linguistics. University of Leiden, the Netherlands.