Khitan Transcriptions Revisited

Marc Hideo Miyake
University of Hawai'i at Mānoa

0. Introduction

Although much progress has been made on the decipherment of the Khitan script in recent years, most efforts have largely ignored the question of what Chinese sounded like during the Liao Dynasty (916-1125 CE). Scholars seem to assume that the Chinese of the time was not unlike Old Mandarin of the Yuan Dynasty. In fact, some scholars go even further by using *pinyin* romanization with only the slightest of modifications to indicate Liao Dynasty pronunciations.

For example, Daniel Kane (1989: 15) writes that the last line in the *Da Jin huangdi dutong jinglüe langjun xingji* inscription dated 1134 CE "can be shown to be read *sh-ang sh-u zhi fang l-ang zh-ung huang ying k-i.*" Other than the substitutions of *u* for *o* and *k* for *q* to reflect the earlier pronunciations of the syllables *zhong* and *qi*, Kane's "phonetic values" could pass for Modern Standard Chinese, a language which obviously did not exist in the 12th century--or in the 10th century, when the Khitan script was devised. Nishida Tatsuo (1980) makes similar anachronistic errors when he discusses the same inscription. I suspect that these errors can be traced to the Chinese researchers whose works Kane and Nishida quote from.

If we are to reconstruct Khitan properly, we must look at Khitan characters--and the Chinese transcriptions of Khitan words in the *Liao shi* and elsewhere--from the perspective of Late Middle Chinese (LMC), not Old Mandarin or modern Mandarin. 'LMC' is a cover term for post-*Qieyun* (601 CE), pre-Song Dynasty standard Chinese. Here, I use LMC to specifically refer to 10th century Chinese.

In this paper, I will (1) present a sketch of my reconstruction of LMC, (2) apply my reconstruction of LMC to the Khitan transcriptions in the *Liao shi* to come up with tentative reconstuctions of the Khitan phonological system and of selected Khitan words from the *Liao shi*, and (3) briefly discuss the implications of this study for future research on the Khitan writing system.

1. The reconstruction of Late Middle Chinese (LMC)

Since this paper is primarily about Khitan and not about LMC, I can only present the barest outline of the methodology and results of my LMC reconstruction. A full justification of my LMC reconstruction is forthcoming. Like Pulleyblank (1984, 1991: 3), I assume that LMC is the language of the *Yunjing* rhyme table, reflecting a phonological tradition traceable to circa the 9th to 10th centuries CE. Many from Karlgren onward have used the *Yunjing* to reconstruct Middle Chinese (as opposed to LMC in particular), but few have focused on hard evidence for LMC such as Tibetan and Turkic transcriptions, Sino-Korean, and Sino-Vietnamese. The reconstruction of LMC presented here relies heavily on such Sinoxenic evidence. Sino-Korean in particular is valuable because it demonstrates how the sounds of LMC can correspond to those of an 'Altaic'-type language; the sounds of LMC may correspond to those of Khitan in a similar way.

My reconstruction of LMC has the following major characteristics:

1. A series of 'muddy' initials which have "voiceless onset with voiced aspiration" (Pulleyblank 1991: 6), derived from the voiced initials of Early Middle Chinese (EMC).

2. A four-way contrast (plain, rhotacized, velarized, palatalized) for [+grave] labial and velar initials, corresponding to the four *deng* of the *Yunjing*.

3. An eight-vowel system.

4. Four tones, including *rusheng*, represented by the lenited codas *-x (*-ɣ?), *-r, *-Φ (*-β?). *-x is attested in Turkic transcriptions and in naturalized Middle Korean loans from Chinese ending in *-h*. *-r is attested in Tibetan and Turkic transcriptions and in Sino-Korean. I know of no attestation of *-Φ except for a single questionable case of *V* in Turkic transcription. All tones but *rusheng* will be ignored in this study.

The following five tables present my reconstruction of LMC in more detail and show how the sounds of LMC correspond to those of my Khitan reconstruction. The LMC-Khitan correspondences will be explained in more detail in section 2.

Table 1. LMC finals, organized by the 16 *she*, the 43 *zhuantu* of the *Yunjing*, and the 4 *deng*

Waizhuan		I	II	III	IV	*Neizhuan*		I	II	III	IV
1. Guo											
果	27.	-ɑ	-	-	-						
	28.	-wɑ	-	-wɑ							
2. Jia						3. Yu					
假	29.	-	-a	-a	-a	遇	11.	-	-o	-o	-o
	30.	-	-wa	-	-		12.	-o	-u	-u	-u
4. Xie						5. Zhi (-i after dentals)					
蟹	13.	-ay	-ay	-ey	-ey	止	4.	-	-i	-i	-i
	14.	-way	-way	-wey	-wey		5.	-	-	-wi	-wi
	15.	-ɑy	-ay	-ey	-ey		6.	-	-i	-i	-i
	16.	-wɑy	-way	-wey	-wey		7.	-	-wi	-wi	-wi
							8.	-	-	-i	-
							9.	-	-	-i	-
							10.	-	-	-wi	-
6. Xiao						7. Liu					
效	25.	-ɑw	-aw	-ew	-ew	流	37.	-əw	-u	-u	-iw
	26.	-	-	-	-ew						
8. Dang						9. Zeng					
宕	31.	-ɑŋ	-waŋ	-ɑŋ	-ɑŋ	曾	42.	-əŋ	-əŋ	-iŋ	-iŋ
	32.	-wɑŋ	-	-wɑŋ	-		43.	-wəŋ	-	-wiŋ	-
10. Geng											
梗	33.	-	-aŋ	-eŋ	-eŋ						
	34.	-	-waŋ	-weŋ	-weŋ						
	35.	-	-aŋ	-eŋ	-eŋ						
	36.	-	-waŋ	-	-weŋ						
11. Jiang						12. Tong					
江	3. -yaŋ					通	1.	-oŋ	-oŋ	-uŋ	-uŋ
							2.	-oŋ	-oŋ	-oŋ	-oŋ
13. Shan						14. Zhen					
山	21.	-	-an	-en	-en	臻	17.	-ən	-ən	-in	-in
	22.	-	-wan	-en	-wen		18.	-on	-	-win	-win
	23.	-ɑn	-an	-en	-en		19.	-	-ən	-in	-
	24.	-wɑn	-wan	-wen	-wen		20.	-	-	-un	-
15. Xian						16. Shen					
咸	39.	-am	-am	-em	-em	深	38.	-	-im	-im	-im
	40.	-ɑm	-am	-em	-em						
	41.	-	-	-em	-						

Table 2. LMC transcription onset-Khitan onset correspondences

EMC initial	LMC '36 initials' by *deng*				Khitan value
	I	II	III	IV	
p	p	p^{r}	$p^{ɥ}$	p^{y}	*b
ph	ph	ph^{r}	$ph^{ɥ}$	ph^{y}	*p
b	pɦ	$pɦ^{r}$	$pɦ^{ɥ}$	$pɦ^{y}$	*p
m	m	m^{r}	$m^{ɥ}$	m^{y}	*m
p, ph	-	-	f	-	no *f in native words?
b	-	-	fɦ	-	no *f in native words?
m	-	-	v	-	no *v in native words?
t	t	-	-	t^{y}	*d
ʈ	-	ʈ	ʈ	-	*d
th	th	-	-	th^{y}	*t
ʈh	-	ʈh	ʈh	-	*t
d	tɦ	-	-	th^{y}	*t
ɖ	-	ʈɦ	ʈɦ	-	*t
n	n	-	-	n^{y}	*n, also *ñ?
ɳ	-	ɳ	ɳ	-	*n, also *ñ?
k	q	q^{r}	$k^{ɥ}$	k^{y}	*g
kh	qh	qh^{r}	$kh^{ɥ}$	kh^{y}	*k (*[q] before back vowels)
g	-	-	$kɦ^{ɥ}$	$kɦ^{y}$	*k (*[q] before back vowels)
ŋ	ŋ	$ŋ^{r}$	$ŋ^{ɥ}$	$ŋ^{y}$	no initial *ŋ in native words?
ts	ts	-	-	ts^{y}	*j
ʈʂ	-	ʈʂ	-	-	*j
c	-	-	c	-	*j
tsh	tsh	-	-	tsh^{y}	*c
ʈʂh	-	ʈʂh	-	-	*c
ch	-	-	ch	-	*c
dz	tsɦ	-	-	$tsɦ^{y}$	*c
ɖʐ	-	ʂɦ	-	-	*ś
j	-	-	śɦ	-	*ś
s	s	-	-	s^{y}	*s
ʂ	-	ʂ	-	-	*ś
ś	-	-	ś	-	*ś
z	-	-	-	$sɦ^{y}$	*s
ʐ	-	ʂ	-	-	*ś
ź	-	-	śɦ	-	*ś
ʔ	ʔ	ʔ	ʔɥ	ʔy	*Ø
h	h	h^{r}	$h^{ɥ}$	h^{y}	*X (may be *x or *ɣ) *q?
ɦ	hɦ	$hɦ^{r}$	-	-	*X (may be *x or *ɣ) *q?
w	-	-	w	-	*w
y	-	-	-	y	*y
l	l	l	$l^{ɥ}$	l^{y}	*l (but LMC *-r + *l- = Khitan *-r-)
ñ	-	-	r	-	*r (and *ź?)

Table 3. LMC surface vowels

	front	central	back
high	i	ɨ	u
mid	e	ə	o
low	a		ɑ

Table 4. LMC transcription vowel-Khitan vowel correspondences

LMC vowel *(kaikou/hekou)*	*Deng* I	II	III	IV	
a, ɑ	a	a	-	-	
wa, wɑ	(w)a	(w)a	-	-	
i	-	-	*ï	*i	
		*I	*I	*I	front/back indeterminate after non-grave initials
wi	-	-	*ü	*ü	(or just *i or *wi?)
ɨ	*ï	*ï	*ï	*ï	
e	-	-	*ö	*e	
	-	-	*E	*E	front/back indeterminate after non-grave initials
we	-	-	*ö	*ö	(or just *e or *we?)
ə	*ö	*ö	*ö	*ö	
wə	*ö	*ö	*ö	*ö	
u	*u	-	*ü	*ü	
		*U	*U	*U	front/back indeterminate after non-grave initials
o	*o	-	*ö	*ö	
		*O	*O	*O	front/back indeterminate after non-grave initials

Table 5. LMC transcription coda-Khitan coda correspondences

LMC coda	Khitan codas
-ŋ	*-ŋ (only in Chinese loans and before velars in native words?)
-x (ɣ?)	*-K (may be *-k, *-x, *-g, *-G; even *-Ø?)
-n	*-n
-r	*-R (may be *-l, *-r; reconstructed as *-r- if preceding *l- initial sinograph)
-m	*-m
-Φ (β?)	*-P (may be *-p, *-b)

2. Towards a tentative reconstruction of Khitan

In preparing this paper, I analyzed the spellings of approximately 100 Khitan words given in Chinese transcription in the *Liao shi*. The spellings had the following characteristics:

* There were no transcription characters with LMC *f-, *ffi-, *v- or *ŋ-.
* Nearly all other LMC initials were represented. All four *dengs* were represented.
* All eight LMC vowels were represented.
* The *rusheng* coda LMC *-r was very frequent, particularly in combination with a following LMC *l-.
* However, the coda LMC *-x was infrequent, and the coda LMC *-Φ was very rare.

From these observations, I conclude:

* Khitan did not have *f, *v, or initial *ŋ-, at least in native words. The frequent appearance of LMC *ph and *pfi shows that Khitan *p had not lenited.
* Scribes may have employed the full range of initials (LMC *q, *q^r, *$k^{ɪ}$, *k^y, for /k/ in the four *dengs*) to represent distinctions between Khitan *q and *k and Khitan front and back vowels.
* The eight vowels, in conjunction with the full range of *dengs*, suggest a complex Khitan vowel system.
* The frequent coda-onset cluster LMC *-r# #l- may be a convention for writing intervocalic Khitan *-r-. If so, intervocalic Khitan *-l- would be written with LMC *l- *not* preceded by LMC *-r. Since *r- generally cannot appear in initial position in Altaic languages, word-initial LMC *l- unambiguously represents Khitan word-initial *l-. However, cases of LMC *-r before initials other than LMC *l- and word-finally remain ambiguous and could represent either Khitan *-r or *-l.
* Since Khitan had no *f or *v, the very rare LMC coda *-Φ (*-β?) must represent Khitan *-p, *-b (or possibly *-x, *-γ?), not *-f, *-v.

On the basis of the transcription data, I reconstruct the following minimal phonemic inventory for Khitan:

Table 6. Khitan consonants

	velar	palatal	dental	labial
-voc stops	*k (*[q],*[k])	*c	*t	*p
+voc stops	*g	*j	*d	*b
nasals	*-ŋ (loans?)	(*ñ?)	*n	*m
-voc fricatives	*x	*ś	*s	(*f??)
+voc fricatives	(*γ?)	(*ź??)		(*v??)
glides and liquids		*y *r	*l	*w

In cases where an LMC consonant could transcribe more than one Khitan consonant, I have listed the less marked consonant (*n, *x) without a question mark and listed more marked possibilities (*ñ, *γ) with single question marks. Double question marks signify consonants that may be in Chinese loanwords which did not appear in the *Liao shi* glossary.

I could easily assume that the eight vowels of LMC are directly equivalent to the eight vowels of Khitan, but for the time being, I will assume instead that Khitan had a more Altaic-looking vowel system:

Table 7. Khitan vowels

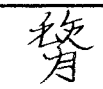
	+front		-front	
	-round	+round	-round	+round
+ high	*i	*ü	*ï	*u
- high	*e = *ä	*ö	*a	*o

*E, *U, *O represent vowels of indeterminate roundness. *I represents a vowel of indeterminate frontness.

The above is not purely typological conjecture. I hypothesize that the four *dengs* and the *kaikou/hekou* distinction were used to distinguish Khitan *i : *ï, *e : *ö, *u : ü, and *o : *ö. See Table 4 for a tentative attempt to show how the LMC vowels were used to represent non-LMC vowels. I believe that the Khitan vowel writing conventions parallelled the strategies used by Old Japanese orthography to distinguish *kô-otsu* vowels. Since no vowel length is reconstructible on the basis of the LMC transcriptions, I will ignore it in my reconstructions. There is apparently no vowel harmony in Khitan.

Using the above tables and the transcriptions in the *Liao shi*, I have reconstructed the following Khitan words. Numbers refer to page numbers in *Liao shi, juan* 116 *"Guoyu jie"* ('The National Language Explained). Reconstructions given in Kane (1989) are listed for comparison whenever available.

Table 8. Sample reconstructed Khitan words

Liao shi transcription	Khitan native spelling	Gloss	Miyake	Kane (1989)
沙里				
116.3; LMC *ṣa li		'prince'	*śalI	*śaali
时				
116.15; LMC *phɑ		'time'	*pa	*po
陶里				
116.15; LMC *tɦɑw li		'hare'	*taulI	*taolia
捏褐	伏为			
116.16; LMC *nyer hɦɑr		'dog'	*NeRXaR	*nehei
抹里	马			
116.18; LMC *mɑy li; probably scribal error for LMC *mɑr li used in *Liao shi, juan* 31		'kind of official' < 'horse'?	*marI	*mor
朱實得本				
116.19; LMC *chex śir təx pon		'filial piety'	*ceXśiRdöbon	*śiistpu

[This list will be expanded to include all Khitan words in the *Liao shi, Liao shi shiyi,* and *Qidan guozhi.*]

3. Conclusion

The above proposal for an LMC-based reconstruction needs further refinement. Khitan native spellings may or may not clarify ambiguities in the above LMC-based reconstruction. A reconstruction of the Northern Song language may also prove to be valuable, for there is no guarantee that all of the transcriptions in the *Liao shi* and elsewhere were meant to be read in LMC--some transcriptions may have been devised in the 10th century (or even earlier in some cases?) while others may have been devised in the 11th and 12th centuries.

Returning back to the *Da jin huangdi dutong jinglüe langjun xingji* inscription of 1134 with which I began this paper, perhaps its last line may be better read:



*ś- -aŋ ś- -ö jIX faŋ laŋ dUŋ huaŋ iŋ q- -ï

Although LMC was dead by the 12th century, the Khitan may have used 'Sino-Khitan' spellings set in the LMC period to write Chinese. Even if this line turns out to be a transcription of Song Dynasty Chinese, I doubt that the scribe ~~did not have~~ had nearly modern "*shang shu zhi fang lang zhung huang ying ki*" in mind.

Research on the Chinese language(s) of the Liao Dynasty must go hand in hand with research on the Khitan language and script. With so few bilingual sources of data extant for Khitan, we have to do the best we can with what little we have. An anachronistic interpretation of the Chinese data is bad enough; building a Khitan reconstruction ultimately based on that anachronistic interpretation is even worse.