Khitan Transcriptions Revisited

Marc Hideo Miyake
EALL 750
16 December 1996

0. Introduction

After the fall of the Tang 唐 Dynasty (618-907) and the chaos of the Five Dynasties and Ten Kingdoms 五代十國, a 鼎 "tripod" (Janhunen 1994: 107) emerged consisting of (1) the Northern Song 北宋 in China proper (960-1127/1127-1279), (2) the Khitan 契丹/ Liao 遼, in southern Mongolia and Manchuria (907-1125), and (3) the Tangut 唐古 / 黨項 / Xixia 西夏 in Gansu 甘肅 (982/1032-1227). Both the non-Sinitic Khitan and Tangut empires had their own scripts which probably served as symbols of independence from Chinese culture. I consider these scripts to be 'quasisinographic': they resemble sinographs superficially but are actually constructed according to radically different principles. Although the Tangut script has supposedly been 'deciphered' twice (Nishida 1964-66, Sofronov 1968) with dubious results, the Khitan script remains largely undeciphered, despite many fragmentary attempts over several decades.

The decipherment of the Khitan script and the reconstruction of the Khitan language have been hampered by a variety of factors. Kane (1989: 11) lists three: "the paucity of the material [in Khitan script], the badly eroded state of the inscriptions and the lack of bilinguals." I would add a fourth: the lack of explicit and sound methodologies for decipherment and reconstruction. In this paper, I will critique previous attempts at decipherment and reconstruction, present my own proposed methodology, and apply the first steps of this methodology to the Khitan words found in Chinese transcription in the *Liao shi* 遼史, the *Liao shi shiyi* 遼史拾遺, and the *Qidan guozhi* 契丹國志.

1. The Khitan writing systems

Before criticizing methodologies of decipherment, a few remarks are needed concerning the objects of decipherment.

The elusive Khitan 'script' is actually two scripts, the Khitan Large Script Small Script 契丹大文 and the Small Script 契丹小文.

The Large Script was created by the decree of the first Liao emperor Taizu 太祖 / *Yälür Abaugï 耶律阿保機 ("Yelü Abaoji"; 907-926) in 920. It consists of approximately 600[1] characters used solely for their semantic values; the closest non-Khitan parallel may be the Japanese *kunyomi* 訓讀み. Each and every character stands for only one morpheme and cannot be used phonetically in rebuses. One must know both the meaning of a Large Script character and the morpheme with that meaning in the Khitan language to read a Large Script character; without both pieces of knowledge, only guesswork is possible.

The semantic basis of the Large Script prevented it from ever being successful. Kane's (1989:10) remarks on the early Jurchen script "based on the Khitan large (ideographic) script", apply equally well to the Khitan Large Script: "The script could be used for enumerating items, but could not express grammatical functions or record words for which a special character had not been invented" (Kane 1989: 10). These limitations prevented the Large Script from ever being a true system of writing, capable of writing any word in the language it represented.

---

[1]I calculated this figure by subtracting the 378 Small Script characters from the total of "1000 characters" (Kane 1989: 11) known to exist in the Khitan scripts. Wittfogel and Feng (1949: 243) quote the *Wudai shiji*'s 五代史記 (72: 4b) figure of "several thousand characters" in the Large Script, but although the extant hundreds of Large Script characters no doubt do not comprise the whole of the Large Script, I severely doubt that there ever were thousands of purely semantic characters, for even the Chinese script has only a hundred or so semantic characters. The claim that the Tangut script is composed of thousands of semantic characters is dubious for the same reason.

However, the Khitan were not without a true system of writing for long. In 925, only five years after the invention of the Large Script, *Yälür Terar? 耶律迭剌 ("Yelü Diela"), younger brother of Emperor *Abaugï, created the Small Script. Unlike the Large Script, the Small Script was entirely phonetic; its symbols stood for sounds or sequences of sounds, not whole morphemes. These symbols were combined into blocks representing whole words, much as *hankul* 한글 letters are combined into blocks representing syllables. The *Liao shi* (64: 5a)[2] tells us that *Terar created the Small Script after studying the Uighur language for twenty days; some conclude from this that *Terar learned of phonetic writing from the Uighurs and applied the Uighurs' alphabetic principle imperfectly to the Khitan Small Script. Although the script's inspirational origins remain uncertain, there is no doubt that the Khitan Small Script was more successful than its immediate predecessor. Only the earliest extant Khitan writings are entirely in Large Script; the rest are in a mixture of Large and Small Scripts. (There are no extant texts entirely in Small Script.) But even this mixture of Large and Small Scripts was not apparently as successful as the Tangut script; writings in the (apparently[3]) more complex Tangut script vastly outnumber those in the Khitan scripts.[4]

---

[2]*Liao shi* citations refer to *juan* 卷, page number, and side of page in the 116-volume *Qianlong wuyingdian* 乾隆武英殿 edition.

[3]My own research on the Tangut script indicates that many of its characters are structurally similar to the Khitan Small Script; they are combinations of phonetic symbols, not semantic radicals, as generally believed. Other Tangut characters are similar to *xingsheng* 形聲 sinographs, combing phonetic symbols with semantic radicals. Thus the Tangut script, contrary to popular belief, may not be composed of thousands of *huiyi* 會意 characters.

[4]However, as Alexander Vovin (class discussion) has pointed out, the Tangut script also was believed to be little used until Kozlov discovered the Tangut library at Qaraqoto. Unfortunately, no Khitan equivalent of the Qaraqoto library has ever been found. Thus one cannot say for sure that the Khitan Large-Small Script hybrid was a failure, though there is little doubt about the failure of the Khitan Large Script.

When the Khitan Empire fell in 1125, the Khitan language disappeared along with it. Remnants of the Khitan Empire founded Qarā Khitāy / *Xi Liao* 西遼, but this successor state in Transoxiana seems to have left no traces of the Khitan language or script. The Khitan script outlived the Khitan state; the Khitans' worst enemies, the Jurchens 女眞, kept the script (and probably also the language) alive until 1191, even after the creation of the Jurchen Large and Small scripts in 1119 and 1138 respectively.

The Khitan scripts were all but entirely forgotten in the centuries that followed. Their existence was noted in the dynastic histories of the Khitan compiled long after the empire's demise, but their appearance was unknown until they were rediscovered (and mistakenly identified as "Jurchen") in the 19th century. The proper identification of Khitan had to wait until the 1920s. Numerous attempts have been made since then to decipher Khitan, but none have ever claimed the success of the so-called 'decipherments' of Tangut. Furthermore, all of these attempts have had methodological problems.

2. Critique of previous attempts at Khitan decipherment and reconstruction

In general, previous attempts at Khitan decipherment and reconstruction (1) lacked an explicit, consistent methodology and (2) had problems with Chinese reconstructions. Too many of the books and articles available to me on Khitan reconstruction[5] employ an approach best described as 'magical'. Reconstructions are simply posited 'out of the blue' *without any justification at all.*

---

[5]I do not claim to have read every single extant work on Khitan. The vast bulk of this work, done in the former USSR and in China, is inaccessible to me. Thus my critique may be unfair. However, I do have second-hand knowledge of Russian and Chinese work on Khitan from the extensive quotations found in those works which are accessible to me.

Nishida Tatsuo's (1980: 112) series of articles on Khitan in *Gekkan gengo* 月刊言語 begins with an example of this school of 'reconstruction'. Judging from reconstructions in this example such as *dörben 'four' (cf. Mong. *dörben* 朶兒邊[6]), the apparent underlying assumption was that the semantic Large Script characters would be read in Middle Mongolian. Why Middle Mongolian? Nishida implies that previous researchers chose Middle Mongolian on the basis of the sinograph 耆 *tshak (actually EMC *ɕa!), which supposedly appears (in the *Liao shi*) as a transcription for the word for 'time', resembles Mong. *caɣ* 'time' 察黑 in the *Yuan chao bi shi* 元朝秘史.[7] On the basis of one supposed cognate (whose true non-*rusheng* reading bears almost no resemblance to Mong. *caɣ*), an *entire sentence* (*dörben caɣ heme jige kürüyal laqu jada[n] cidaqu, glossed sinograph-by-sinograph as 四時節到來再來不能) was reconstructed from Large Script characters!

As Nishida correctly points out, such an approach lacks concrete supporting evidence. Even if Khitan were a relative of Middle Mongolian, there is absolutely no guarantee that any word in its lexicon would be identical or similar to a word of the same meaning in Middle Mongolian. After all, dialects of a language—and, it goes without saying, related languages separated by centuries such as Khitan and Middle Mongolian—do not have 100% identical lexicons. Practitioners of this school of reconstruction equate their assumptions (that Khitan was a relative of Middle Mongolian) with reality, shoving Khitan into a Middle Mongolian mold with barely any justification.

The temptation to do this is understandable, though not justifiable, for the Large Script characters are the hardest to crack. Some words written in Large Script also appear in Chinese

---

[6]All Mongolian forms are quoted with *Hua-yi yiyu* 華夷譯語 transcriptions when available from Mostaert (1977), *Le matérial mongol du* houa i i iu 華夷譯語 *de houng-ou* (1389).

[7]See below for a discussion of the problems with this claim.

transcriptions and in Small Script, so their readings can be reconstructed, but what is one to do with those words in Large Script which never appear in any other script? Assuming those words were identical to Middle Mongolian words is an easy (and dangerous) 'solution'.

'Magical' reconstructions are hardly limited to the Large Script. Even the Small Script has been the object of speculation.

Tamura and Kobayashi's (1952-3) Keiryō 慶陵 comes with an appendix of charts of Small Script characters and reconstructed readings. Although the appendix is valuable as a tabulation of the various attested combinations of Small Script characters, the reconstructions are dubious because the authors give no justification whatsoever for the reconstructed readings, either in the appendix or in the text of Keiryō itself.

Chen Naixiong 陳乃雄 (1988) gives a list of Large and Small Script characters with reconstructed readings from a variety of Chinese works on Khitan, again with no justification whatsoever except for a handful of Mongolian words. The reconstructions are particularly suspicious as they (1) make extremely fine Chinese-style distinctions (between *s, *ʂ, and *ʃ, for example), (2) are very un-Altaic-looking, with a large number of vowels (but no *ö or *ü!) and retroflex consonants, and (3) give multiple readings for a single character. None of the above characteristics are necessarily impossible. Khitan may have lacked *ö and *ü like Manchu. It could have had an extensive loan phonology to accomodate 'Sino-Khitan'.[8] And Occam's Razor aside, there is no

[8]The existence of Sino-Khitan is dubious, mainly because there is no evidence that the Khitan were as heavily Sinicized as the Vietnamese or the Koreans. While other states in the Sinosphere used Chinese titles and wrote in Classical Chinese, the Khitans used Khitan titles and made bilingual inscriptions. Furthermore, if the Koreans, who were even more heavily Sinicized than the Khitans, did not import Chinese phonology wholesale, complete with retroflexes and all, then the probability of the Khitans doing so is very small.

reason to believe that each character had one and only one reading. However, extraordinary claims require extraordinary justification. None of the summaries of Chinese research on Khitan available to me give such extraordinary justification. Apparently, the characters are read in a certain way because...they just are. The point here is not to say that the reconstructions are necessarily wrong, but that there is no way to know whether they are right.

Not all Khitan reconstructions are necessarily 'magical'. Others, in particular those involving Chinese transcriptions and the Small Script, are quite rooted in concrete evidence...and, unfortunately, questionable Chinese reconstructions. None of the works on Khitan known to me ever addressed the fundamental question: What was the Chinese of the tenth century CE like?

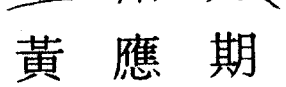
Those works that reconstruct Khitan via reconstructions of the pronunciations of Chinese transcriptions of Khitan used (a) Karlgrenian reconstructions of Early Middle Chinese (EMC; 7th century CE), (b) 'antiqued' Mandarin without palatalization of velars, etc., (c) modern Mandarin in non-phonetic Wade-Giles or Pinyin romanization (!), or (d) a combination of any of the above (!!). None used a reconstruction of Northeastern Late Middle Chinese (NELMC; hereafter, simply LMC) as spoken during the three centuries of Liao rule (907-1125).

Karl Menges' study of the Khitan language in Wittfogel and Feng (1949: 429-34) used both Karlgren's EMC reconstruction and *non*-'antiqued' modern Mandarin readings (!) to decipher Chinese transcriptions of Khitan words. Wittfogel and Feng note that

> ...sharp discrepancies appear in many instances between [Karlgren's reconstruction of] Sui-T'ang Chinese and Altaic forms. Frequently the final consonant of Sui Chinese finds no correspondence in early or modern Altaic, whereas the modern Chinese forms, because of their loss of the final consonants (except *-n* and *-ng*), do not present this difficulty...Examples of this type suggest that the post-T'ang Chinese phonology as reflected in the Liao transcription of Altaic words had gone a long way in abandoning the final consonants characteristic of earlier Chinese.

This claim allowed Menges to intepret the Chinese transcriptions of Khitan words in any form of Chinese—either EMC or modern Mandarin—that would facilitate an alleged Altaic etymology. For example, Menges (Wittfogel and Feng 1949: 430) used the modern Mandarin reading *zha* for LMC *kaf 閘 (cf. Sino-Korean *kap*, Sino-Vietnamese *sáp* < *krap) to claim that the Khitan term LMC *kaf.sar 閘撒 might have been a transcription of Mongol *jasaq* 札撒 'law'. On the very same page, Menges used Karlgren's reconstruction *xi̯wət 狘 (LMC *hwyeR) to claim that the Khitan term LMC *kaf.sar.hwyer 閘撒狘 was cognate to Mongol *jasa'nut* (plural of *jasaq*). If Menges is correct, then the transcription 閘撒狘 should be read with the modern Mandarin initial *zh-* and the EMC coda *-t*—a combination of features that could never have existed in the same period. There is no consistency or realism in this approach. Without an explicit agenda, anything is possible.

Several decades later, anachronism still reigns in Khitan reconstruction. The Khitan Script Research Group's recent reconstruction of the last line from the *Da jin huangdi dutong jinglüe langjun xingji* 大金皇弟都城經略郎君行記 inscription of 1134, cited both by Kane (1989: 14) and Nishida (1980: 116) without critical comment, strongly resembles modern Mandarin, except for the use of *-u-* for *-o-* and *k-* for *q-*:

Khitan					
Chinese	尚	書	職	方	郎	中	黃	應	期
Mandarin	*shang*	*shu*	*zhi*	*fang*	*lang*	*zhong*	*huang*	*ying*	*qi*
K.S.R.G.	*sh- -ang	sh- -u	zhi	fang	l- -ang	zh- -ung	huang	ying	k- -i
Miyake	*ʃ- -ang	ʃ- -ö	jIK	fang	l- -ang	di/j- -ung	huang	ïng	q- -ï

If one takes the Pinyin-like reconstruction at face value, then one might suppose that the characters 尚書職中 were all pronounced with retroflex fricative and affricate initials (Pinyin *sh-*, *zh-*) in the early 12th century. This might have been a possibility, for Pulleyblank (1984) claims that 尚書職 were already pronounced with retroflex initials in LMC. However, in any case, the use of Pinyin in reconstructions suggests that researchers simply equated the Chinese of Liao times with 'antiqued' Mandarin, despite the centuries-wide gulf between the two.

One might have expected at least some researchers to have used reconstructions of Yuan Dynasty Chinese to interpret the Chinese transcriptions of Khitan. Since the transcriptions are found in histories of the Khitan compiled during the Yuan Dynasty, there is a possibility that the transcriptions may reflect Yuan Dynasty phonology rather than Liao-era phonology. Unfortunately, the works I have seen devoted as much space to Yuan Dynasty reconstructed Chinese as to LMC—which is to say no space at all. Only the extremes of Karlgren EMC and modern Mandarin ('antiqued' or not) were represented. This paper may be the very first study of Khitan that takes LMC into account.

3. The present methodology outlined

As a Sinologist, I intend to tackle the problem of reconstructing Khitan from the Chinese side for two reasons. First, I lack the necessary training required to reconstruct Khitan from the Mongolic side.[9] Second and more importantly, sinographic transcriptions of Khitan and Khitan transcriptions of sinograph readings are the most solid evidence at present for Khitan, since they correlate the

[9]The precise Altaic affiliations of Khitan have long been controversial. Khitan has been variously identified as Turkic, Mongolic, Tungusic, or an Altaic *Mischsprache*. On the basis of the vocabulary I have examined, I am inclined to identify it as Mongolic.

known (Chinese) with the unknown (Khitan). Working solely within the unknown (Khitan) without any reference to the known (Chinese) would be highly risky, since there would be no constraints (i.e. sinograph readings) to prevent one from simply shoving Khitan into a Middle Mongolian mold.[10]

Admittedly, the extensive use of Chinese data has risks of its own. There is no guarantee that Chinese transcriptions are 100% accurate. Indeed, the *Liao shi shi yi* has corrections to the spellings in the *Liao shi*. And even the 'correct' spellings might not be able to express sounds that may not have been in the Chinese of the time, such as uvulars or front rounded vowels. The transcriptions present Khitan 'with a Chinese accent', not Khitan itself—the gulf between the two might be as wide as that between *katakana*-ized English and the 'real thing'. Furthermore, the orthographies of the *Yuan chao bi shi* and the Mongol section of the *Hua yi yiyu* have diacritics which are entirely lacking in the Chinese transcriptions of Khitan. The absence of such diacritics makes one doubt the precision of the Chinese data. Nevertheless, it's all we've got. Not surprisingly, the very first step of my strategy for reconstructing Khitan involves Chinese:

> *1. Reconstruct Northeastern Late Middle Chinese (LMC)—the variety of Chinese known to the Khitan between the tenth and early twelfth centuries CE.*
>
> This first step has only been partly accomplished; I will present the initial results in section 4. Sections 5 and 6 of this paper present the results of the second through fifth steps, still in progress as of this writing:

---

[10]The 'reconstructions' of Tangut by Nishida (1964-66) and Sofronov (1968) were faulty precisely because they emphasized internal evidence (the *Tangraphic Sea* or *Wenhai* 文海) over external sinographic and Tibetan alphabetic evidence. Without the constraints of Chinese and Tibetan transcriptions, Nishida and Sofronov were free to shove Tangut into the desired Tibeto-Burman mold, just as earlier Khitan researchers shoved Khitan into one Altaic mold or another.

2. Convert Chinese transcriptions of Khitan from the Liao shi, Liao shi shiyi, Qidan guozhi *and other sources into LMC phonetic strings.*

3. Uncover the outlines of Khitan phonology, based on the inventory and distribution of the LMC phonemes represented in the transcriptions.

4. Create skeletal reconstructions of Khitan words.

5. Use these skeletal reconstructions to test previously proposed etymologies.

Section 7, the conclusion of this paper, briefly describes the last steps which will be performed in later paper(s):

6. Fill in skeletal reconstructions using cognates in known Altaic language(s).

7. Use outside cognates to sort out the strata (if any) of transcriptions.

8. Work on post-LMC Chinese and Old Mandarin to deal with the latest strata.

9. Revise reconstructed pronunciations of late strata transcriptions accordingly.

10. Test reconstructions with attested Small Script spellings.

11. Finally assign reconstructed sound values to Large and Small Script characters.

12. Begin work on the Jurchen script, an offshoot of the Khitan script.

4. The reconstruction of Northeastern Late Middle Chinese (LMC)

At present, I do not claim to have 'perfected' my reconstruction of LMC. I have yet to undertake an analysis of the *fanqie* 反切 in the character dictionary *Longkan shoujian* 龍龕手鑑, compiled c. 997 by a Khitan monk. I am probably not alone in having failed to do this; I know of no Khitan researcher who has used the *Longkan shoujian* to assist in the interpretation of sinographic transcriptions. But even if I did analyze the *Longkan shoujian,* it would only give me insight into the

categories of initials and finals of Liao-era LMC as opposed to concrete sound values. For such sound values, I have turned to both alphabetic (Tibetan and Turkic) and comparative (Sinoxenic) evidence to produce the tentative reconstruction outlined below. Since this paper is 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 will appear in my PhD dissertation.

I assume that LMC is similar to the language of the *Yunjing* 韻鏡 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 (i.e. a language with relatively few consonants); 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' (全)濁 initials which have "voiceless onset with voiced aspiration" (Pulleyblank 1991: 6), derived from the voiced initials of Early Middle Chinese (EMC).

The devoicing of the 'muddy' (全)濁 initials is clear from Kan-on 漢音, based on a northwestern variety of LMC predating Liao-era LMC by two centuries. Tibetan transcriptions of northwestern Five Dynasties Chinese (Luo 1933), which have devoiced fricatives but voiced stops, would seem to argue against this claim. However, devoicing is also justifiable on the basis of apparent cases of 'muddy' (全)濁 and 'second clear' 次清 initials being used to write the same Khitan word—'muddy'-initial LMC *pɦo 步 and 'second clear'-initial LMC *pho 普 both appear

as the second syllable of the Khitan word *mapo? 馬步 / 麻普 'tribal title'. (Of course, alternate explanations for the doublet are possible—one or the other may be a scribal error or both may represent the same word in two different transcription systems from two different eras, one pre- and one post-devoicing.)

2. A four-way contrast (plain, rhotacized, velarized, palatalized) for [+grave] labial and velar initials, corresponding respectively to the four *deng*s 四等 of the *Yunjing*. This interpretation of the four *deng*s —particularly the second and third—is admittedly rather unusual, and I cannot give anything approaching a complete defense of it here.

The notion of a second *deng* with rhotacized initials was inspired by the fact that the retroflex initials are exclusive to the 2nd *deng* in Middle Chinese. I do not put much faith in this view of the second *deng;* its initials may have merely been velarized like those of the third. (See below for an alternate view.)

The notion of a third *deng* with velarized initials was inspired by the fact that Sino-Korean has *-u-* and *-e-* in third *deng* rhymes contrasting with *-i/y-* and *-ye-* in fourth *deng* rhymes. Sino-Korean represented the contrast between LMC *k^{ɯ}i (*k^{ɯ}ɨ?; *deng* III) and *k^{y}i (*deng* IV) as *kuy* (*[kɯi] or *[kɨy]?) 긔 and *ki* (*[ki]) 기.

I formerly advocated a view of the four *deng*s which I called the 'А-Я-Ы-И hypothesis' (via correlating reconstructed LMC vowels [and palatalizations of preceding consonants] with equivalent Russian letters):

deng	initial	vowel		vowel class
I:	non-palatalized	back	(*ɑ, *o, *u)	"А, О, У"
II:	palatalized	low front	(*a)	"Я"
III:	non-palatalized or velarized	high~mid central~front	(*ɨ, *e)	"Ы, Э"
IV:	palatalized	high~mid	(*i, *e, *o, *u)	"И, Е, Ё, Ю"

Although now obsolete, this system and its successor both have velarized initials (or should it be velar semivowels?) for the third *deng* contrasting with palatalized initials (or a palatal semivowel) for the fourth *deng*. This hypothesis feeds my hypothesis about how the sinographic transcriptions correlated with Khitan vowels, as I will explain below.

3. An eight-vowel system, which frequently correlates directly with Sino-Korean (given here in Yale romanization) and Sino-Vietnamese (given here in phonetic transcription, not *quôc ngu'* orthography):

LMC	SK	SV	LMC	SK	SV	LMC	SK	SV	LMC	SK	SV
*a	*a*	a:	*i	*i*	i, ə	*ɨ	*u, o*	ɨ, a	*u	*wu*	u
*ɑ	*a*	a:	*e	*e*	e	*ə	*o, a*	ə, a	*o	*wo*	o

4. Four tones, including *rusheng* 入聲, represented by the lenited codas *-x (*-ɣ?), *-r, *-Φ (*-β?). *-x is attested in Turkic transcriptions as *-X* 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 (*LĀV* 臘; Barat n.d.: 3).

All tones but *rusheng* will be ignored in this study. The non-*rusheng* tones appear to be interchangeable in variant transcriptions of the same word. For example, the previously cited spellings of *mapo? 馬步 / 麻普 'tribal title' have characters of different tonal classes (*shang* 上 and *qu* 去 in the case of the former spelling, *ping* 平 and *shang* 上 in the case of the latter).

The following five tables on pages 15-17 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 5.

LMC finals

Organized by the 16 *she* 攝, the 43 *zhuantu* 轉圖 of the *Yunjing*, and the 4 *deng*

Waizhuan 外轉	I	II	III	IV	*Neizhuan* 內轉	I	II	III	IV
1. Guo 果									
27.	-ɑ	-	-	-					
28.	-wɑ	-	-wɑ						
2. Jia 假					3. Yu 遇				
29.	-	-a	-a	-a	11.	-	-o	-o	-o
30.	-	-wa	-	-	12.	-o	-u	-u	-u
4. Xie 蟹					5. Zhi 止 (-ɨ after dentals)				
13.	-ay	-ay	-ey	-ey	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 效					7. Liu 流				
25.	-ɑw	-aw	-ew	-ew	37.	-əw	-u	-u	-iw
26.	-	-	-	-ew					
8. Dang 宕					9. Zeng 曾				
31.	-ɑŋ	-waŋ	-ɑŋ	-ɑŋ	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 江					12. Tong 通				
3. -yaŋ					1.	-oŋ	-oŋ	-uŋ	-uŋ
					2.	-oŋ	-oŋ	-oŋ	-oŋ
13. Shan 山					14. Zhen 臻				
21.	-	-an	-en	-en	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 咸					16. Shen 深				
39.	-am	-am	-em	-em	38.	-	-im	-im	-im
40.	-ɑm	-am	-em	-em					
41.	-	-	-em	-					

Possible 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
PU, PHU	-	-	f	-	no *f in native words?
BU	-	-	fɦ	-	no *f in native words?
MU	-	-	v	-	no *v in native words?
T	t	-	-	t^{y}	*d
TR	-	ʈ	ʈ	-	*d
TH	th	-	-	th^{y}	*t
THR	-	ʈh	ʈh	-	*t
D	tɦ	-	-	th^{y}	*t
DR	-	ʈɦ	ʈɦ	-	*t
N	n	-	-	n^{y}	*n, also *ɲ?
NR	-	ɳ	ɳ	-	*n, also *ɲ?
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)
NG	ŋ	$ŋ^{r}$	$ŋ^{ɥ}$	$ŋ^{y}$	no initial *ŋ in native words?
TS	ts	-	-	ts^{y}	*j
TSR	-	ʈʂ	-	-	*j
C	-	-	c	-	*j
TSH	tsh	-	-	tsh^{y}	*c
TSHR	-	ʈʂh	-	-	*c
CH	-	-	ch	-	*c
DZ	tsɦ	-	-	$tsɦ^{y}$	*c
DZR	-	ʂɦ	-	-	*ʃ
J	-	-	ʃɦ	-	*ʃ
S	s	-	-	s^{y}	*s
SR	-	ʂ	-	-	*ʃ
SY	-	-	ʃ	-	*ʃ
Z	-	-	-	$sɦ^{y}$	*s
ZR	-	ʂ	-	-	*ʃ
ZY	-	-	ʃɦ	-	*ʃ
Q	ʔ	$ʔ^{r}$	ʔɥ	ʔy	*Ø
H	h	h^{r}	$h^{ɥ}$	h^{y}	*H (may be *x, *ɣ, or even *q)
X	hɦ	$hɦ^{r}$	-	-	*H (may be *x, *ɣ, or even *q)

Possible LMC transcription onset-Khitan onset correspondences (continued)

EMC initial	LMC '36 initials' by *deng*				Khitan value
	I	II	III	IV	
V	-	-	w	-	*w
Y	-	-	-	y	*y
L	l	l	$l^{(l)}$	l^{y}	*l (but LMC *-r + *l- = Khitan *-r-)
NY	-	-	r	-	*r (and *ʒ?)

LMC vowels

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

Possible LMC transcription vowel-Khitan vowel correspondences

LMC vowel (*kaikou* 開口 and *hekou* 合口)	*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 [-grave] initials
wi	-	-	*ü	*ü	(or just *i or *wi?)
ɨ	*ï	*ï	*ï	*ï	
e	-	-	*ö	*e	
	-	-	*E	*E	front/back indeterminate after [-grave] initials
we	-	-	*ö	*ö	(or just *e or *we?)
ɘ	*ö	*ö	*ö	*ö	
wɘ	*ö	*ö	*ö	*ö	
u	*u	-	*ü	*ü	
		*U	*U	*U	front/back indeterminate after [-grave] initials
o	*o	-	*ö	*ö	
		*O	*O	*O	front/back indeterminate after [-grave] initials

Possible 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, *-ɣ; even *-Ø?)
-n	*-n
-r	*-R (may be *-t, *-d, *-l, *-r; reconstructed as *-r- if preceding *l-)
-m	*-m
-Φ (β?)	*-P (may be *-p, *-b, even *-Ø?)

5. Correspondences between LMC and Khitan

Assuming that my reconstruction of LMC is correct, how does one then use that reconstruction in turn to reconstruct Khitan? An extreme and absurd approach would be to use the LMC reconstruction as is, ensuring that no information in the transcription would be lost (and that no information outside the transcription would be added). The sole advantage of such an approach would be the prevention of reconstructions deliberately altered to resemble alleged Altaic cognates.

My approach, on the other hand, sticks closely to the LMC reconstruction while also leaving some leeway for minor alterations. In short, I have mapped the known sounds of LMC onto the hypothesized sounds of Khitan.

These hypothesized sounds have their basis in both the transcriptional data and general Altaic (and in particular, Mongolic) phonological typology. First, I analyzed the spellings of approximately 150 Khitan words in Chinese transcription in the *Liao shi* 遼史, the *Liao shi shiyi* 遼史拾遺, and the *Qidan guozhi* 契丹國志. The spellings had the following characteristics:

1. There were very few transcription characters with LMC *f-, *ffi-, or *v-.

2. There were only three words with initial *ŋ-.

3. Nearly all other LMC initials were represented. All four *dengs* were represented.

4. All eight LMC vowels were represented.

5. Words may contain 'mixed' sequences of vowels: front and back, rounded and unrounded.

6. The *rusheng* coda LMC *-r was very frequent, particularly when followed by a sinograph with initial LMC *l-.

7. However, the coda LMC *-x was infrequent, and the coda LMC *-Φ was very rare.

From these observations, I concluded:

1. Khitan probably did not have *f or *v in native vocabulary.

2. The frequent appearance of LMC *ph and *pɦ as well as the absence of labiodental initials shows that Khitan *p had not lenited to *f.

3. Initial *ŋ- may have been in loans only. The two words with this initial, *ŋaʃu 牙署 / 牙書 'tribal title' and *ŋualI 瓦里 'tribal unit' contain the character LMC *ŋ^{r}a 牙 'fang' or its near homonym LMC *ŋ^{w}a 瓦 'tile'. There is a remote possibility that the first word may be a borrowing from some Chinese original meaning something like 'fang scribe'. Another remote possibility is that the spellings 牙署 / 牙書 reflect bogus etymologies projected onto the word by sinographic scribes. If this were the case, then the word might not have even begun with *ŋ- at all—it might have had some similar initial like *n- or *g-. However, neither of these highly unlikely explanations can account for *ŋualI 瓦里 'tribal unit'.

4. Scribes may have employed the full range of initials (LMC *q, *q^{r}, *k^{ɯ}, *k^{y}, for /k/ in the four *deng*s) to represent distinctions between Khitan *q and *k and Khitan front and back vowels.

5. The eight vowels, in conjunction with the full range of *deng*s, suggest a complex Khitan vowel system.

6. The cooccurrences of vowels belonging to different classes (front vs. back; rounded vs. unrounded) within words suggest that Khitan may have had no vowel harmony.

7. The frequent coda-onset cluster LMC *-r# #l- may be a convention for writing intervocalic Khitan *-r-. Some (but not all) words with apparent Altaic cognates containing *-r- such as *ɲEri 捏離 'sun' (cf. Mong. *nara(n)* 納古蘭 'sun') appear with LMC *-r# #l-, confirming this hypothesis. If this hypothesis is correct, then 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- would unambiguously represent Khitan word-initial *l-. However, cases of LMC *-r (a)

before initials other than LMC *l- and (b) word-finally remain ambiguous and could represent either Khitan *-r or *-l.

8. 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 above facts and hypotheses, and with Altaic typology in mind (no retroflex or aspirate stop series, for example), I reconstructed the following minimal consonantal inventory:

Possible Khitan consonants

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

In cases where an LMC consonant could transcribe more than one possible Khitan consonant, I have listed the less marked consonant (*n, *h/x) without a question mark and listed more marked (but not necessarily improbable) possibilities (*ɲ, *γ) with single question marks. Double question marks signify consonants that may be in Chinese loanwords which did not appear in the *Liao shi, Liao shi shiyi,* or *Qidan guozhi.*

The table on page 16 shows how these proposed Khitan consonants would be represented by LMC consonants in transcriptions. I have assumed that LMC voiceless aspirated : Khitan voiceless and LMC voiceless unaspirated : Khitan voiced, on the model of the pattern found in Chinese transcriptions of Mongolian, Jurchen, and Manchu. This could turn out to be utterly wrong, particularly if it turns out that Khitan had a three-way distinction between, say, *p : *ph : *b. But that doesn't seem typologically likely, given the lack of distinctive aspirate series in Mongolian and Manchu. It may, however, be possible, that LMC aspirates sometimes represented Khitan fricatives

(i.e. LMC *qh : Khitan *x). But Khitan didn't seem to have any labial fricatives, so how would one explain the occurrences of LMC *ph in the transcriptions?

Words like *ɲeRHaR (*ɲeRqaR?) 捏褐 'dog' (cf. Mong. *noqai* 那孩 'dog') suggest the opposite of the above proposed LMC *aspirate : Khitan *fricative pattern. In some cases, the pattern of correspondence may have been LMC *fricative : Khitan *stop. Khitan syllable-finals may have been written with LMC final fricatives. (See the chart of coda correspondences on page 17.)

I have assumed that Khitan backed */k/ to *[q] before back vowels, for two reasons. The first is typological (cf. other 'Altaic'-type languages); the second is the fact that the transcription sinographs used to write Sanskrit velar initial-back vowel combinations such as LMC *kfiʷa 伽 were never used in Khitan transcription, suggesting that Khitan did not have *ka; one finds 可 LMC *qhɑ instead, presumably for Khitan *qa in words like *qaton (*qatun?) 可敦 'wife of a qaɣan'. But since the Sanskrit transcription sinographs were probably 'off limits' for writing non-Sanskrit, particularly 'barbarian' languages, the absence of sinographs like 伽 may actually signify nothing.

Did Khitan have an backed allophone of */g/ before back vowels? The presence of sinographs like 古 LMC *qo (Khitan *go [ɢo]?) with uvular initials and the absence of near-homophones with velar initials like 居 LMC *kʷo (which would have represented Khitan *go [go]) in transcriptions seems to indicate allophonic backing of Khitan */g/ (*[ɢ]?).

The retroflex stops probably already affricated by the tenth century, although they are still reflected as stops in Sino-Korean (currently palatals in modern Korean pronunciation) and in Sino-Vietnamese (*tr* in *quôc ngu'*). Late Tang and Five Dynasties Tibetan transcriptions have palatals, not *-r-* clusters, for these initials. Hence LMC retroflex stops *ʈ, *ʈh, *ʈfi should probably be rewritten as LMC *ʈʂ, *ʈʂh, *ʈʂfi, corresponding to Khitan *j, *c, *c, not Khitan *d, *t, *t.

Unlike Chinese researchers, I seriously doubt that Khitan had a three-way ts-ʈʂ-c distinction in native words or even in loan phonology. (See note 8 on page 6 for my arguments against 'Sino-Khitan'.) Hence, I conflated the LMC affricates, considering them all to either represent Khitan *c or *j on the basis of the presence or absence of aspiration. When analyzing the Khitan writing system, however, one should still be on the lookout for any graphemes making a partial or total distinction between dental, retroflex, and palatal affricates, especially in Small Script compound graphs representing Chinese loans.

LMC *h and *hfi could be rewritten as *x and *xfi; there is no hard evidence to reconstruct these two phonemes as either glottals or velars. Khitan might have had both an *h and an *x (or just one and not the other) corresponding to LMC *h and *hfi or even *kh and *kfi. But without a definite reconstruction of the phonetic value of LMC *h, the sinographic transcriptions alone cannot clarify this issue. The presence of *h* and the absence of *x* in the Mongolian of the *Yuan chao bishi* and *Hua-yi yiyu* tells us nothing, for Mongolian *h*, unlike Khitan *H, derives from earlier *p-, which never lenited in Khitan, judging from the presence of LMC *ph- and *pɦ- and the absence of LMC *f- in transcriptions.

Compared to the many-to-one correspondences between LMC and possible Khitan consonants, the one-to-one vocalic correspondences are vastly simpler. I could easily assume that the eight LMC vowels are directly equivalent to eight Khitan vowels of Khitan, but for the time being, I will assume instead that Khitan had a more Altaic-looking vowel system:

Possible Khitan vowels

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

The above is not purely typological conjecture. I hypothesize that the four *deng*s and the *kaikou/hekou* distinction found in the transcriptions were used to partly distinguish Khitan *i : *ï, *e : *ö, *u : ü, and *o : *ö:

—*deng* III LMC *i with velarized initials corresponds to Khitan *ï (central or back unrounded).

—LMC *ɵ and *deng* III LMC *e with velarized initials (*[-^{ɯ}ə-]; Sino-Korean *-e-*) correspond to Khitan *ö (front mid rounded) due to the acoustic similarity of unrounded back or central vowels with front rounded vowels.

—*deng* III LMC *u with velarized initials corresponds to Khitan *ü (front high rounded) for the acoustic reasons given above.

—Since velarization in *deng* III applies only to [+grave] initials and since a preceding velarized initial is the key to determining whether an LMC *i corresponds to Khitan *i or *ï, LMC *i after [-grave] initials in *deng* III corresponds to indeterminate Khitan *I (*i or *ï?). Similar logic motivates reconstructing *E (*e or *ö?), *U (*u or *ü?), *O (*o or *ö?) after [-grave] initials.

—*deng* IV LMC *o, *u with palatalized initials correspond to Khitan *ö, *ü.

Since no vowel length is reconstructible on the basis of the LMC transcriptions, I will ignore it in my reconstructions for now. External comparisons will probably make the reconstruction of Khitan vowel length (if it existed at all) possible.

I will put aside the question of how to interpret LMC vowel-glide sequences like *-iw which might correspond to Khitan *-iu, *-iw, or even *-ü. Tentatively, I interpret LMC vowel-glide sequences as transcriptions of Khitan vowel-vowel sequences.

The reconstructions of LMC and Khitan in particular presented thus far are incomplete and unsatisfying. Many points remain questionable if not unknown. However, recall that thus far I have only progressed to step two of the twelve-step agenda that I set out on pages 10-11. The results presented on the previous pages are not the end product, but only the beginning.

The results of step three—the application of my LMC and Khitan reconstructions to the transcriptions of Khitan words in the *Liao shi, Liao shi shiyi,* and the *Qidan guozhi*—are presented in the appendix.

Sample results of step four—the testing of previously proposed reconstructions and etymologies using the current LMC and Khitan reconstructions—are presented in the next section.

6. Re-evaluations of Khitan-Altaic reconstructions and etymologies

Due to limitations of space, I will only present a handful of comparisons between earlier reconstructions and mine. Numbers refer to *juan* and page of the *Liao shi.*

Liao shi transcription	Khitan native spelling	Gloss	Miyake	Kane (1989)
抹里 31.2; LMC *mɑr li	 (Large Script)	'kind of official' < 'horse'?	*marI	*mor
沙里 116.3; LMC *ʂa li	 (Small Script)	'prince'	*ʃalI	*ʃaali
 116.15; LMC *phɑ	 (Large Script)	'time'	*pa	*po
陶里 116.15; LMC *tɦɑw li	 (Small Script)	'hare'	*taulI	*taolia
捏褐 116.16; LMC *nyer hɦɑr	 (Small Script)	'dog'	*ɲeRHaR	*nehei
赤實得本 116.19; LMC *chex ʃir tɘx pon	 (Small Script)	'filial piety'	*ceKʃiRtöKbon	*ʃiistpu

Despite the wide methodological gulf between the modern Mandarin-based reconstructions of the Chinese researchers quoted by Kane (1989) and my LMC-based methodology, the results on both sides are in general not very different from each other. In fact, the reconstructions for 'prince' are identical, apart from vowel length (which Kane derived from the presence of the Small Script graph *a 月). The same could be said of the reconstructions for 'hare'; Kane's extra *-a once again was derived from the presence of the Small Script graph *a月. If my methodology took account of Small Script spellings (as it will at a later stage), then it too would have reconstructed vowel length in 'prince' and the final -a of 'hare'.

One key difference between Mandarin-based reconstructions and LMC-based reconstructions is the vocalism, for both languages are at different stages of what I have termed the 'Great Chinese Vowel Shift'. Examples are 'horse' and 'time', which have *a in my reconstructions corresponding to *o in Kane's. Although the vocalism of 'time' is difficult to argue about because there is no Mong. *ho or *ha 'time' cognate to Khitan *po/*pa, external evidence may favor the more traditional reconstruction with *-o- for 'horse'. Kane's *mor is closer to Mong. *morin* than my *mar. This word and others such as 'ordos' LMC *wɑrlodwɑ 斡魯朶 which would read 'better' if *o or *ɔ were substituted for my *a, indicate that some, if not all, of the transcriptions date after LMC *ɑ > late LMC *ɔ (as represented by the 'Tang Min 唐閩' loan stratum in Min).

There is no reason to believe that all of the transcriptions represent a uniform system, as I have implicity assumed for the sake of simplicity until now. Some spellings surely must date back to Liao times (or even earlier) while others may have been invented on the spot by the Yuan Dynasty authors of the Liao histories. (However, if the latter turns out to be true, then how could the compilers of the histories of the Khitan have heard the Khitan words that they were transcribing

tremendous phonetic attrition. Thus the loss of vowels that would result in clusters like *-stp- (cf. loss of vowels and the resulting clusters in Middle Korean) would not be entirely unexpected in Khitan. Nevertheless, the problem of trying to make *ʃiistpu fit with the transcription 赤實得本 remains; at least my competing reconstruction, *ceKʃIRdöKbon (perhaps simply *ciʃidöbun if *rusheng* endings are dropped and the vowels are shifted to a late LMC/Old Mandarin stage) fits the transcription precisely. Of course, fitting the transcription is no guarantee of correctness. The more I apply the LMC-based method of reconstruction, the less confident I am of my results.

7. Conclusion

Despite any failures at this early stage of my research, I can confidently say that at least I have brought up a valid issue by asking the simple question: What was the Chinese language like in Liao times? My current reconstruction of LMC may turn out to be almost as anachronistic as the Karlgrenian EMC or the modern Mandarin ('antiqued' or not) used by others due to its insufficient vowel shifts and retention of *rusheng* codas, but my future work with the Liao character dictionary *Longkan shoujian* and with the skeletal rhyme table in the Song Dynasty work *Huangji jingshi shu* 皇極經世書 may ultimately save my approach. Once I reconstruct a revised LMC, the cycle will begin anew, with a reanalysis of the transcriptions in terms of the latest set of LMC values.

Even if I somehow manage to 'perfect' my LMC reconstruction and sort out all the strata of transcriptions, there will still be many steps left to take before the decipherment process comes to an end. I will need to study Turkic, Mongolic, and Tungusic in order to (1) judge the validity of previous etymologies and, if possible, (2) discover new etymologies of my own. Working purely from the Chinese side in ignorance of Altaic prevents me from arbitrarily skewing the reconstructions

Appendix: The Chinese transcriptions of Khitan vocabulary in the *Liao shi, Liao shi shiyi,* and *Qidan guozhi*

This list comprises nearly all Khitan words found by the author in the *Guoyu jie* 國語解 section of the *Liao shi* and *Liao shi shiyi* and the *Suishi zaji* 歲時雜記 section of the *Qidan guozhi.* It is *not* a complete list of all Khitan words found in all three works; some variant spellings and many proper names not in the *Guoyu jie* or *Suishi zaji* have been omitted. Eventually, those words will be added and any 'Khitan' words found to be of Chinese origin will be omitted.

Due to typographical difficulties, rare characters will be represented in the form PHONETIC$^{\text{RADICAL}}$ or $^{\text{RADICAL}}$PHONETIC, depending on the position of the phonetic. For example, 寧^{頁} represents a single character with the phonetic 寧 on the left and the radical 頁 on the right.

All forms given (EMC, LMC, Khitan) are reconstructions given *without* asterisks.

EMC forms are cited in the conventionalized orthography partly outlined in my handout "Go-on and Pre-Old Japanese Phonology". The symbols -Ø, -Q, and -H at the end of a syllable represent *pingsheng, shangsheng,* and *qusheng* respectively. Medial -R- represents velarization, not *[-r-]. The digraph UI represents *[ɨ]. V represents *[w]. W represents *[ɯ]. X represents *[ɦ].

Due to typographical difficulties, LMC forms are cited without superscripts in an orthography similar to that used for EMC. All tones but *rusheng* have been omitted.

Due to typographical difficulties, Khitan reconstructions are given in a simplified orthography, with capital letters substituting for *ʃ and the umlauted vowels and the digraphs ny and ng substituting for ɲ and ŋ. Indeterminate vowels have not been marked.

Glosses are generally from Wittfogel and Fêng (1949); any exceptions are my translations from the Chinese glosses.

Abbreviations in citations:

LS	=	*Liao shi*
LSSY	=	*Liao shi shiyi*
QDGZ	=	*Qidan guozhi*
WF	=	Wittfogel and Fêng (1949)

Citations for the first three works follow the format JUAN.PAGE. Thus "LS 116.11" means *Liao shi, juan* 116, page 11. Note that the pagination of the *LSSY* and *QDGZ* are cumulative; page numbers are those of the entire work, not of the individual *juan.*

Words are indexed by EMC readings in alphabetical order.

Transcription	EMC	LMC	Khitan reconstruction	Gloss	Citation
婆離	BAA. LUIQ	pxaa. li	pali	tribal name	LS116. 11
暴里	BAAVH. LUIQ	pxaaw. li	pauli	name of villain	LS116. 4
倍其不離	BAYQ. GUI. BUT. LIA	pxaay. kxwi. pxur. li	paikIpuri	kind of drum	LSSY24. 435
蒲割寧[貫]	BO. KAT. NENYQ	pxo. kaar. nyeng	pogaRneng	tribal title	LS116. 7
蒲速盌	BO. SOK. QVAANQ	pxo. sox. qvaan	posoKwan	prosperous	LS116. 18
皮室	BRIA. SYIT	pxri. syir	pISiR	name of army	LS116. 17
別輦斗	BRIAT. LIANQ. TAVQ	pxryer. lyen. tav	perendOu	toponym	LS116. 11
赤實得本	CHIAC. ZYIT. TAK. PONQ	chyex. sxyir. tax. pon	ceKSirtOKbon	filial piety	LSSY23. 428
炒伍侕尉	CHRAAV. NGOQ. NYUI	chraav. ngo. ri	caungori	fight	LS116. 16
敞史	CHUIANGQ. SRUIQ	chwang. sri	cOngSi	tribal title	LS116. 8
敞穩	CHUIANGQ. QONQ	chwang. qon	cOngon	tribal title	LS116. 10
俎里	CRUIA. LUIQ	crwo. li	jOli	tribal title	WF432:LS63. 7
達剌干	DAAT. LAAT. KAAN	txaar. laar. kaan	taraRkan	tribal title	WF33:LS46. 1-2
達剌子	DAAT. TSHIAH. TSUIQ	txaar. tshi. tsi	taRciji	tribal title	LS116. 7
陶里	DAAV. LUIQ	txaav. li	tauli	hare	LS116. 15
大迭烈特	DAAYH. DET. LIAT. DAK	txaay. txyer. lyer. txax	taitereRtOK	Diela territory	LS116. 2
特滿	DAK. MAAN	txax. maan	tOKman	many	WF429:LS46. 13, 19
提烈	DE. LIAT	txyey. lyer	teleR	seeding fields	LSSY 24. 434
題里	DE. LUIQ	txyey. li	teli	tribal title	LS116. 24
迪輦	DEC. LIANQ	txyec. lyen	teKlen	tribal title	WF432:LS63. 7
敵烈痲都	DEC. LIAT. MRAA. TO	txyec. lyer. mraa. to	teKleRmado	tribal title	LS116. 14
徒覩古	DO. TOQ. KOQ	txo. to. ko	todogo	tribal title	LS116. 8
獨盧金	DOK. LO. KIM	txox. lo. kim	tologim	toponym	LS116. 11
治夔離	DRIH. GRVI. LIA	txri. kxrvi. li	ciqili	*wansui!*	LSSY24. 434
墮瑰	DVAAQ. KOY	txvaa. koy	tuagoi	name of gate	LS116. 20
奪里本	DVAAT. LUIQ. PON	txvaar. li. pon	tuarIbon	subdue	LS116. 18
旗鞁拽剌	GUI. KOQ. YIAYH. TSHIAH	kxi. ko. yey. tshi	kigoyeci	tribal title	LS116. 14
虎番[去]	HOQ. PHAN	ho. phaan	hopan	personal name	LS116. 11
虎斯	HOQ. SIA	ho. si	hosi	personal name	LS116. 25
虎思	HOQ. SUIQ	ho. si	hosi	personal name	LS116. 13
忽兒珊	HOT. NYIA. SAAN	hor. ri. saan	horisan	title of general	LS116. 13
懽撻新查剌	HVAAN. THAAT. SIH. JRAA. TSHIAH	hvaan. thaar. si. cxraa. tshi	huantaRsicaci	LS116. 13	
石烈	JIAC. LIAT	cxyex. lyer	ceKReR	subtribe	LS116. 2, 16, 22

Transcription	EMC	LMC	Khitan reconstruction	Gloss	Citation
常衮	JUIANG. KONQ	cxwang. kon	cOnggon	tribal title	LS116. 5
閘撒狨	KAAP. SAAT. HUAT	kaaf. saar. hwyer	gaPsaRheR	official?	LS116. 8
葛兒罕	KAAT. NYIA. HAANQ	kaar. ri. haan	garihan	gurxan	LS116. 13
高墩	KAAV. TON	kaav. ton	gaudon	tribal title?	LS116. 12
可敦	KHAA. TON	khaa. ton	qaton	Turk empress	LS116. 23
剋	KHAK	khak	qak	military unity	LS116. 6, 24
契丹	KHIT. TAAN	khir. taan	kiRdan	Khitan	-
起兒漫	KHUIQ. NYIA. MAANH	khi. ri. maan	kiriman	toponym	LS116. 13
孤穩	KO. QONQ	ko. qon	goon	jade	LS116. 18
楞楜遲	KOT. TSVAAT. DRI	kor. tsvaar. txri	goRjaRji	thousand-yr snake	LS116. 24
監母	KRAAM. MAVQ	kraam. mav	gammau	leave behind	LS116. 19
君基太乙	KUN. KUI. THAAYH. QIT	kun. ki. thaay. qir	gungitaiiR	deity name	LS116. 5
國阿輦	KVAK. QAA. LIANQ	kvax. qaa. lyen	guaKalen	? country (Ch. loan)	LS116. 18
落昆髓	LAAK. KON. SVIAQ	laax. kon. svi	laKgonsui	toponym	LS116. 13
令穩	LIANYH. QONQ	lyeng. qon	lengon	official title	LS116. 16
林牙	LIM. NGRAA	lim. ngraa	limnga	official title	LS116. 7
莫弗賀	MAAK. BUT. XAAH	maax. fur. hxaa	maKfurha	title of chief	WF428:LS63. 2
莫弗紇	MAAK. BUT. XAT	maax. fur. hxar	maKfurhaR	title of chief	LS116. 22
莫^{邑}頡	MAAK. XET	maax. hxyer	maKheR	administration name	LS116. 1
抹鶻	MAAT. XOT	maar. hxor	maRhoR	tribal title	LS116. 17
抹里	MAAT. LUIQ	maar. li	mari	tribal title	LS116. 18
梅里	MAY. LUIQ	may. li	mayli	tribal title	LS116. 17
麻都不	MRAA. TO. PU(T)	mraa. to. pu(r)	madobu(R)	tribal title	LS116. 8, 17
麻普	MRAA. PHOQ	mraa. pho	mapo	tribal title	LS116. 16
馬特本	MRAAQ. DAK. PONQ	mraa. txax. pon	matabon	tribal title	LS116. 16
馬步	MRAAQ. BOH	mraa. pxo	mapo	tribal title	LS116. 8
彌里	MYIA. LUIQ	myi. li	mili	administrative unit	LS116. 2, 16
南撒葛栢	NAM. SAAT. KAAT. PRAAC	naam. saar. kaar. praax	namsaRkaRpaK	toponym	LS116. 11
迺捏咿喨	NAYQ. NET. QYI. NYIA	naay. nyer. qyi. ri	nainyeiri	head day?	LS116. 15
耨斡麼	NAVQ. VAAT. MAA	nav. vaar. maa	nauwarma	empress	LS116. 23
捏骨地	NET. KOT. DIH	nyer. kor. txi	nyeRkoRti	kneel	LSSY24. 435
捏離	NET. LUIQ	nyer. li	nyeri	day	QDGZ27. 250
捏褐耐	NET. XAAT. NAYH	nyer. hxaar. naay	nyeRhaRnai	dog head	LS116. 16
牙署	NGRAA. JUIA	ngraa. cxi	ngaci	tribal title	LS116. 8

Transcription	EMC	LMC	Khitan reconstruction	Gloss	Citation
瓦里	NGRAA. LUIQ	ngraa. li	ngali	tribal title	LS116. 17
牙書	NGRAA. SYUIA	ngraa. syi	ngasi	tribal title	LS116. 8
訛莎烈	NGVAA. SRAA. LIAT	ngvaa. sraa. lyer	nguaSaleR	toponym	LS116. 12
女古	NRUIAQ. KOQ	nrwo. ko	nOgo	gold	LS116. 18
女古底	NRUIAQ. KOQ. TEQ	nrwo. ko. tyey	nOgode	toponym(gold. . . ?)	LS116. 13
妳	NUIQ	ni	ni	4th Heavenly Stem?	QDGZ27. 250
濃兀	NYUAVNG. NGOT	nyvong. ngor	nyongngoR	tribal unit name	LS116. 9
時	PHAAQ	phaa	pa	time	LS116. 15
頗里	PHAAQ. LUIQ	phaa. li	pali	tribal name	LS116. 11
方裉朵殿	PUANG. QYIN.	faang. qyin. tvaa	fangyindua	kind of palace	LS116. 15
必里遲離	PYIT. LUIQ. DRI. LIA	pyir. li. txri. li	pilicili	9th mo. 9th day	LS116. 16
阿札割只	QAA. CRAAT. KAAT. CIAQ	qaa. craar. kaar. ci	ajaRgaRji	tribal title	LS116. 10
阿主沙里	QAA. CUAQ. SRAA. LUIQ	qaa. cu. sraa. li	ajuSali	name	LS116. 3
阿魯敦	QAA. LOQ. TON	qaa. lo. ton	alodon	?	WF431:LS73. 3
阿盧朵里	QAA. LUIA. TVAAQ. LUIQ	qaa. li. tvaa. li	aliduali	?	LS116. 5
阿里軫斗	QAA. LUIQ. CINQ. TAVQ	qaa. li. cin. tav	alijindau	toponym	LS116. 13
溫里謹	QAV. LUIQ. KUINQ	qaa. li. kuin	aligIn	toponym	LS116. 13
阿斯	QAA. SIA	qaa. si	asi	vast	LS116. 19
阿厮	QAA. SIA	qaa. si	asi	vast	LSSY23. 428
阿思	QAA. SUI	qaa. si	asi	vast	LSSY23. 428
阿點夷离的	QAA. TEMQ. YI. LIA. TEC	qaa. tyem. yi. li. tyex	ademyilideK	wife of official	LS116. 4
諲譔	QEN. TSHVIAN	qyen. tshvyen	encuen	ruler of Bohai	LS116. 6
烏魯古阿里只	QO. LOQ. KOQ. QAA. LUIQ. CIAQ	qo. lo. ko qaa. li. ci	ologo aliji	horse name	LS116. 6
嗢娘改	QOT. NYUIANG. KAYQ	qor. nywang. kay	oRnyOnggai	toponym	LS116. 3
乙室板里	QRIT. SYIT. PRAANQ. LUIQ	qrir. syir. praan. li	IRSiRpanli	clan names	LS116. 23
窩篤盌	QVAA. TOK. QVAANQ	qvaa. tox. qvaan	uadowan	caring?	LS116. 19
撒里乃	SAAT. LUIQ. NAYQ	saar. li. naay	sarinai	toponym	LS116. 12
撒剌	SAAT. TSHIAH	saar. tshi	saRci	wine goblet	LS116. 23
賽伊唲耆	SAYH. QYI. NYIA. SYAA	saay. qyi. ri. syaa	saiyiriSa	month good	LS116. 16
先離撻覽	SEN. LUIQ. THAT. LAAMQ	syen. li. thar. laam	Senlitaram	Xi and Bohai title	LS116. 17
蕭	SEV	syev	Seu	surname	LS116. 1
辛袞	SIN. KON	sin. kon	singon	tribal title	LS116. 16
沙里	SRAA. LUIQ	sraa. li	Sali	prince	LS116. 2

Transcription	EMC	LMC	Khitan reconstruction	Gloss	Citation
思奴古	SUI. NO. KOQ	si. no. ko	sinogo	tribal title	LS116. 8
算斡魯朶	SVAANQ. VAAT. LOQ. TVAAQ	svaan. vaar. lo. dvaa	suanwarodua	heart belly ordo	LS116. 18
選底	SVIANQ. TEQ	svyen. tyey	Sende	prison official	LS116. 5
赦例	SYAAH. LIAYH	syaa. lyey	Sale	office	LSSY24. 434
舍利	SYAAQ. LIH	syaa. li	Sali	office	LS116. 5
世燭	SYIAYH. CUK	syey. cux	SejuK	tribal title	LS116. 8
得失得本(誤)	TAK. SYIT. TAK. PONQ	tax. syir. tax. pon	dOKSiRdOKbon	filial piety	LS116. 19
戴辣	TAYH. LAAT	taay. laar	dailaR	burn armor	LS116. 16
柢柘	TEQ. XOH	tyey. hxo	teho	gate guardian?	LS116. 24
拖古烈	THAA. KOQ. LIAT	thaa. ko. lyer	tagoleR	toponym	LS116. 11
塔里捨	THAAP. LUIQ. SYAAQ	thaaf. li. syaa	taPliSa	toponym	LS116. 12
撻馬(狘)	THAAT. MRAAQ. HUAT	thaar. mraa. (hwyer)	taRma(heR)	escort(s?)	LS116. 2, 9
撻林	THAAT. LIM	thaar. lim	tarim	tribal title	LS116. 5
闥林	THAAT. LIM	thaar. lim	tarim	tribal title	WF434:LS 46. 23-24
討賽咿哂	THAAVH. SAYH. QYI. NYIA	thaav. saay. qyi. ri	tausaiyiri	five months	LS116. 15
腻俚寋	THAK. LUIQ. KUIANQ	thax. li. kwan	tOkligOn	empress	WF431:LS71. 1
忒里蹇	THAK. LUIQ. KUIANQ	thax. li. kwan	tOKligOn	empress	LS116. 23
梯里己	THE. LUIQ. KUIQ	thyey. li. kwi	teligI	tribal title	LS116. 7
惕隱	THEC. QUINQ	thyex. qwin	teKIn	tribal title	LS116. 3
吐里	THOQ. LUIQ	tho. li	toli	tribal title	LS116. 24
禿里	THOK. LUIQ	thox. li	toKli	tribal title	WF432:LS46. 3
楚古	TSHRUIAQ. KOQ	tshri. ko	cigo	tribal title	LS116. 22
且虙	TSUIA. LUIA	tswo. lwo	jOlO	toponym	LS116. 19
斡魯朶	VAAT. LOQ. TVAAQ	vaar. lo. tvaa	warotua	ordo	LS116. 13
于越	VUA. VUAT	vyu. vyer	wiuweR	tribal title	LS116. 3
何魯盌	XAA. LOQ. QVAANQ	hxaa. lo. qvaan	halowan	assist	LS116. 19
曷里狘	XAAT. LUIQ. HUAT	hxaar. li. hwyer	hariheR	toponym	LS116. 12
曷蘇館	XAAT. SO. KVAANH	hxaar. so. kvaan	haRsokuan	Jurchen tribe	LS116. 10
合蘇袞	XAP. SO. KONQ	hxaaf. so. kon	haPsogon	Jurchen tribe	LS116. 10
合只忽里	XAP. CIAQ. HOT. LUIQ	hxaaf. ci. hor. li	haPjihori	toponym	LS116. 11
侯里古	XAV. LUIQ. KOQ	hxav. li. ko	hOuligo	toponym	LS116. 12
奚	XE	hxyey	he	tribal name	LS116. 3
奚首	XE. SYUQ	hxyey. syu	heSU	cantonment	LS116. 14
鶻里	XOT. LUIQ	hxor. li	hori	thief	LSSY24. 437

Transcription	EMC	LMC	Khitan reconstruction	Gloss	Citation
霞瀨益	XRAA. LAAYH. QYIAC	hxraa. laay. qyi	halaii	tribal name	LS116. 2
心甲里	XRAAP. LUIQ	hxraaf. li	haPRi	name of date	LS116. 15
崤里	XRAAV. LUIQ	hxraav. li	hauli	Shiwei tribe name	LS116. 5
樺	XVRAAH	hxvraa	hua	hunt	LS116. 15
耶魯盌	YAA. LOQ. QVAANQ	yaa. lo. qvaan	yalowan	prosperous	LS116. 18
耶律	YAA. LVIT	yaa. lvit	yalUt	surname	LS116. 1
夷离堇	YI. LIA. KUINQ	yi. li. kwin	yiligIn	tribal title	LS116. 2, 25
夷离畢	YI. LIA. PYIT	yi. li. pyir	yilipiR	tribal title	LS116. 4
遙輦	YIAV. LIANQ	yev. lyen	yeulen	family name	LS116. 6, 23, 24
曜辣	YIAVH. LAAT	yev. laar	yeulaR	sceptre	LSSY24. 434
習	ZIP	sxif	siP	tribal name	LS116. 3
俟斤	ZRUIQ. KUIN	sxi. kwin	sigIn	Turkic title	LS116. 22
詳穩	ZUIANG. QONQ	sxwang. qon	sOngon	tribal title	LS116. 7
象吻	ZUIANGQ. MUN	sxwang. vun	sOngun	tribal title	LS116. 17
射	ZYAAH	sxyaa	Sa	ask	LS116. 13
實里薛袞	ZYIT. LUIQ. PEC. KON	sxyir. li. pyex. kon	SiripeKgon	cap?	LS116. 20

Bibliography

Barat, Kahar. n.d. "A Turkic Chinese transcription system." Unpublished manuscript.

Chen Naixiong 陳乃雄. 1988. *Qidan zi yanjiu shulüe* 契丹字研究述略. In Eguchi, Paul K., et al, eds., *Languages and history in East Asia: festschrift for Tatsuo Nishida on the occasion of his 60th birthday.* Kyoto: Shokado.

Janhunen, Juha. 1994. "On the formation of Sinitic scripts in mediaeval northern China." *Journal de la Société Finno-Ougrienne* 85:107-124.

Kane, Daniel. 1989. *The Sino-Jurchen vocabulary of the Bureau of Interpreters.* Indiana University Uralic and Altaic Series 153. Bloomington: Indiana University.

Lo Changpei 羅常培. 1933. *Tang wudai xibei fangyin* 唐五代西北方音. Shanghai: Academia Sinica.

Mostaert, Antoine. 1977. *Le matérial mongol du* houa i i iu 華夷譯語 *de houng-ou (1389).* Bruxelles: Institut Belge des Hautes Études Chinoises.

Nishida Tatsuo 西田龍雄. 1964-66. *Seikago no kenkyū* 西夏語の研究. Tokyo: Zauhō.

————. 1980. *Kittan moji—sono kaidoku no shintenkai* 契丹文字—その解讀の新展開. *Gengo* 言語 9.1: 112-119, 9.2.106-115, 9.3: 109-116.

Pulleyblank, Edwin G. 1984. *Middle Chinese.* Vancouver: University of British Columbia.

Sofronov, M. V. 1968. *Grammatika tangutskogo jazyka.* 2 vols. Moskva: Nauka.

Tamura Jitsuzō 田村實造 and Kobayashi Yukio 小林行雄. 1952-53. *Keiryō* 慶陵. 2 vols. Kyoto: Kyoto University.

Tuo Tuo 脫脫. *Liao shi yibaishiliu juan* 遼史一百十六卷. *Qianlong wuyingdian kan ben* 乾隆武英殿刊本. n.d.

Wittfogel, Karl A. and Fêng Chia-shêng. 1949. *History of Chinese society: Liao (907-1125).* Transactions of the American Philosophical Society 36. New York: Macmillan.

Ye Longli 葉隆禮. *Qidan guozhi* 契丹國史. Shanghai guji chubanshe 上海古籍出版社 ed. 1985.

Ye Longli 葉隆禮 and Li E 厲鶚. *Liao shi shiyi* 遼史拾遺. Commercial Press 商務印書館 ed. 1936.