Proposal to encode the Old Turkic ligature ORKHON CI

Anshuman Pandey

pandey@umich.edu
pandey.github.io/unicode

February 17, 2019

1 Proposed character

Glyph	Codepoint	Character name
⅄	10C49	OLD TURKIC LIGATURE ORKHON CI

2 Description

The character ⅄ represents the syllable [či]. In this proposal, it is transliterated as $\widehat{ci}$, where the tie indicates that the value is represented by a single character. The ⅄ is attested in the Old Turkic inscription of Tonyukuk, in the word *kältäčimiz*, which occurs in the seventh line on the south-facing side of the first pillar. The full line is given below (rotated 90 degrees counter-clockwise from the original vertical orientation; see fig. 1 for inscription), and the ligature is indicated by the arrow:

:)ᚺᛕᚺᚱᚴ᛬ᛃᚺᛯᛎᛃᚲ᛬ᚺᛁᛐᚺᛁ᛬ᚺᛃ᛬ᚺᛁᛯ᛬ᛃᛁᛥᛵᛐᛁ᛬ᛃᛲᛃᛇᛐᛐ᛬ᛃᛁᛃᛯ᛬ᛃᛁ᛬ᛮᛁ᛬ᛃᛁᚾᛐ᛬ᛃᛲᛦᛐᛃ᛬ᛃᛃᛐᛧ᛬ᛯᛐᛵᛐ᛬ᚱᛵᛃᛃ᛬ᛵᛃᚺᛃ᛬ᛯᚺᛃᛃᛃ᛬ᚺᛵᛃᚺᛃ

A digitization of the text using the Unicode encoding and representative glyphs for Old Turkic reads as follows, with the addition of ⅄ (in red):

:ᛃᛐᛵᛃᛐ :ᛃᛦᚴᛵᚺ :ᛃᛐᛵᚺ :ᛃᛤᛀᛇᛒ :ᛃᛦᛨᛧ :ᛃᛤᛤᛕᚱᛐ :ᛃᛵᛵᚾ :ᛁᚴᛐᛃᛒᛦᛇᚺᚾ :ᛵᛤᛃᚴᛒᛦᛧᛦ
ᚴᚺᚾᚺᚾᛃᛃ :ᛃᚾᛃᚴᛵᛃ :ᚺᚴ⅄ᚺᛦᛧ :ᚺᚴᚾᛁ :ᛵᛃᛥᛐᚾᛃᛧ :ᛃᛤᚺᛦᛧᛐ

yoγⁿ bolsᵃr : üzgᵘlük̆ : ᵃlp ᵃrmˡš : öŋrä : qii̯tᵃñda : bˡrˡjä : tᵃbγᵃčda : qurˡja : qurˡdˡn̲ta : jir̲ˡja :
oγᵘzda : ˡki üč biŋ : sümᵘz : kᵃltᵃcimˡz : bᵃr mu nä : ˡnčä ötᵘntˡm :

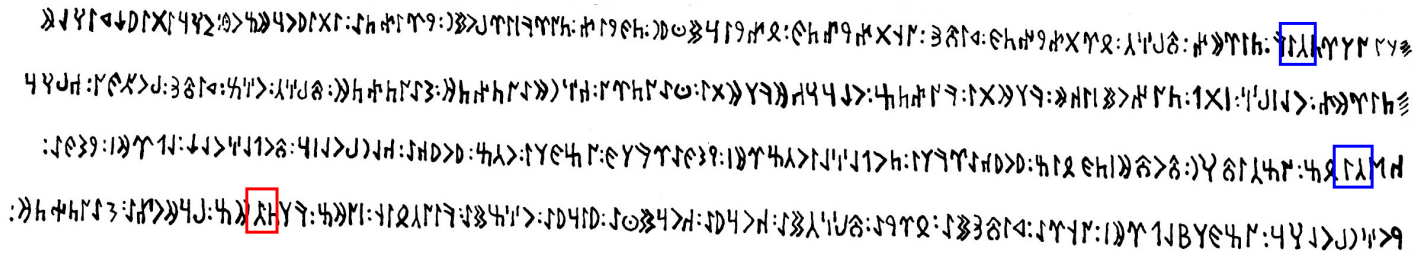The character 𐰯 is a ligature of 𐰲 U+10C32 OLD TURKIC LETTER ORKHON EC + 𐀃 U+10C03 OLD TURKIC LETTER ORKHON I, in which 𐀃 is incorporated into the vertical stroke of 𐰲. The 𐰯 *ci* occurs simultaneously with 𐀃𐰲 *ci͡*, which is used throughout the inscription for the normative representation of [či]. The occurrence of both 𐰯 and 𐀃𐰲 in the Tonyukuk inscription (lines 11–14, 1b south) is shown below:

[Old Turkic inscription — four lines of runiform text]

The conventional representation of [či] occurs in ⸬𐰯𐰞𐰺𐱅𐰖 öl^ürt^äči^ök: (line 11) and ⸬𐰴𐰞𐱅𐰲𐰃𐰕 q^alt^äčib^iz: (line 13), and the ligature in ⸬𐰴𐰞𐱅𐰯𐰢𐰃𐰕 k^alt^ä*ci*m^iz:. While 𐰯 has the same value as the sequence 𐀃𐰲, it is graphically distinct from the latter.

The 𐰯 is similar in appearance to 𐰐 U+10C10 OLD TURKIC LETTER YENISEI AEG, which is used for representing [γ] in the Yenisei form of the script. The two letters differ in the shape of the head. In 𐰯, the right arm connects at the top of the descender, as in 𐀃 ORKHON I, while in 𐰐 the right arm connects to another appendage that extends from a shortened descender. In any case, the morphological contexts of 𐰯 indicate that it is certainly a ligature of 𐰲 + 𐀃, and not the usage of 𐰐 as a proxy for [či].

The 𐰯 is a graphemic *hapax*. This is significant because ligatures are not attested in the Old Turkic script. For reasons that will likely remain unknown, the engraver of the Tonyukuk inscription intentionally introduced an innovation for the script by abbreviating the sequence 𐰲 + 𐀃 as 𐰯. Such scribal practices generally occur at the end of line or where there is insufficient space on the medium. But, this is not the case with *kältäčimiz*, which occurs at some distance from the end of line. The ligature may have been an effort to correct an omission. It is plausible that the engraver had truly intended to produce ⸬𐰴𐰞𐰲𐰞𐱅𐰖, but went from etching 𐰲 to 𐰞, forgetting the intervening 𐀃 (eg. ⸬𐰴𐰞𐰲𐱅𐰖*), and saw an opportunity in the complementary descenders of the two letters to fix the error using ⸬𐰴𐰞𐰯𐱅𐰖. Whatever the reason may be, the ligation of these two letters into 𐰯 signifies a scribal intentionality that departs from the orthographic conventions of the script, but one that is nonetheless unique and distinctive.

The 𐰯 is considered a unique grapheme of the Old Turkic script by scholars (see fig. 4). The character is required for accurately encoding the text of Old Turkic inscriptions, and especially for preserving contrastive representations of [či] (𐰯 and 𐀃𐰲). Therefore, it is necessary to encode 𐰯 as a new character in the 'Old Turkic' block in Unicode.

3 Character Data

Character Properties   In the format of `UnicodeData.txt`:

```
10C49;OLD TURKIC LIGATURE ORKHON CI;Lo;0;R;;;;;N;;;;;
```

Linebreaking   In the format of `LineBreak.txt`:

```
10C49;AL  # Lo    OLD TURKIC LIGATURE ORKHON CI
```

4   References

Sprengling, M. 1939. "Tonyukuk's Epitaph: An Old Turkish Masterpiece. Introduction, Text, Annotated Scientific Translation, Literary Translation and Transliteration". *The American Journal of Semitic Languages and Literatures*, vol. 56, no. 1 (January), pp. 1–19.

Tekin, Talât. 1968. *A Grammar of Orkhon Turkic*. Bloomington: Indiana University / The Hague: Mouton & Co.

———. 2003. *Orhon Türkçesi Grameri*. Türk dilleri araştırmaları dizisi: 9. Yayımlayan: Mehmet Ölmez. Ankara: Türk Dil Kurumu Yayınları.

5   Acknowledgments

Figure 1: Reproduction of the southern face of pillar 1 of the Tonyukuk inscription (from Sprengling 1939). The arrow shows the ligature ci, in the seventh line (indicated as line '(14)').

Thus, all uniconsonantal and vocalic signs of the Turkic alphabet can be explained partly as reproductions of the letters of some Aramaic/Iranian alphabet and partly being the inventions of an inventor; but, this assumption still leaves quite a few, almost a dozen, signs unexplained. These are the so-called 'di-graphs' or 'ligatures'. It has been suggested that these signs, too, were invented by the inventor who may have got this idea from the ligatures in Greek cursive script.[11] The fact that there is no resemblance between any of the 'digraphs' and the letters representing their constituent sounds refutes this theory. The signs for the sound combinations lt, nč, nt, oq/uq/qo/ qu, ŏk/ük/kŏ/kü, ïq/qï, ič, aš, baš, up, and ot cannot be regarded as ligatures, because none of them seems to be a combination of two letters.[12] There is no doubt that these signs are syllabic, not alphabetic. Therefore, it would be sensible to assume that they have an independent origin and developed from ideograms. It should be noted that even Thomsen, founder of the Aramaic origin theory, admitted that the letters y[1] (back y), oq/uq and b[2] might have been ideographic in character, e.g., ay= moon, oq= arrow and ãb= house.[13]

Another difficulty in accepting the Aramaic/Iranian origin theory is that in Aramaic a given sign designates syllables consisting of a given consonant and any vowel, e.g., beth means ba, bi, bu, etc., while in the Turkic

─────────────

[11] Thomsen, Inscrip., pp. 50, 51; G. Clauson, op. cit., p. 74.

[12] The only ligature used in the Orkhon inscriptions is ʎ = či (see 1.2224). This ligature is obviously a combination of the letters ʎ = č and ʌ = i.

[13] Thomsen, L'Alphabet runiforme turc, in: Afh., p. 78.

28

Figure 2: Note regarding c_i as the only attested ligature in the Orkhon inscriptions (from Tekin 1968: 28).

q̊iy¹n²γ q̇ïyïnïγ (TI N 8); ḋiz q̇ïz (generally); ḋizp q̇ïzïp
(TII W 5); q̇ïzl¹ q̇ïzïl (TII E 2).

1.22234. Use of the Syllabic Sign ič

The syllabic sign ič is rarely used to designate the
initial sound group ič: ičgr²t²m ičgärtim (BK E 25);
ičkd²i ičikdi (BK E 37; TI N 4), ičkmš ičikmiš (BK E
9), but generally ičk- ičik-; ič²a ičrä (TI N 10), but
generally ičr²a.

1.22235. Use of the Syllabic Sign aš

The syllabic sign aš is used once in the inscription
of Tonyukuk: ašg ašig? (TI S 1).

I.22236. Use of the Syllabic Sign baš

The syllabic sign baš is used only once in the inscrip-
tion of Tonyukuk: baš (TI N 2).

1.22237. Use of the Ligature λ (= či)

In the following word the medial sound group či is
designated with the ligature λ : kl²t²čimz kältäčimiz
(TI S 7).

1.2224. Use of Compound Consonant-Characters

1.22241. Use of the Sign lt

The compound consonant-character lt is used to re-
present the sound groups alt and lt in back-vocalic words:
    1°. alt: ltm altïm (O R 1), but generally l¹t¹m,
l¹t¹mz, l¹t¹ i; ltzd¹i alt(u)zdï (KT E 38); qltči qaltači
(KT N 9; TI S 6), but ql¹t¹m qaltïm (BK E 14);
    2°. lt: b¹ultči boltači (generally), but b¹ul¹t¹či (KT
N 11; BK E 31); b¹ulti boltï (generally), but b¹ul¹(t¹i)
(KČ W 3), b¹ul¹t¹q̇d¹a (O F 1, 6), b¹ul¹t¹q̇inta (KT N

45

Figure 3: Description of the usage and occurrence of the ligature *c̦i* (from Tekin 1968: 45).

Orhon Yazıtlarında kullanılan işaretlerin tablosu

Harfler	Yazıçevrimi	Ses Değerleri
ᚳ	A	a, ä
ᚱ	I	ı, i
ᚤ	U	o, u
N	Ü	ö, ü
ᛷ (T, KÇ ᛷ)	B	b
ᚸ ᛞ (T, O, KÇ ᛞ)	b	b
ᛝ (KÇ ᛝ)	D	d
✕	d	d
ᛉ (T ᛁ¹)	G	g
ᛇ (O ᛇ)	g	g
ᚼ	K	k
ᚽ (T, O ᚤ, KÇ ᛉ)	k	k
ᚃ	L	l
Y	l	l
᚛	N	n
ᚽ	n	n
ᚣ	R	r
ᛏ	r	r
ᚤ (T ᚤ, KÇ ᛝ)	S	s
I	s	s
ᛘ (T, O, KÇ ᛘ, O ᛞ)	T	t
ᚻ	t	t
ᛈ	Y	y
ᛩ	y	y
ᚶ	ç	ç
ᛣ (T ᛝ)	m	m

Harfler	Yazıçevrimi	Ses Değerleri
ᚻ	ŋ	ŋ
ᚼ	ñ	ñ
ᚠ	p	p
¥	Ş	ş
ᛈ (T, O, ᚽ)	Z	z
↓	UK, KU	ok / uk, ko / ku
ᚱ (T, O ᛒ, KÇ ᛒ)	Ük, kÜ	ök / ük, kö / kü
◀ (T, O ▶)	ıK, Kı	ık / kı
Y	iç, çi	iç
ᚶ	çi	çi
M	LT	lt
ᛷ	nç	nç
☺ (T ☺, O ⊙, KÇ ⊙)	nt	nt
□	aş	aş
⋈	baş	baş

Figure 4: Chart of Orkhon letters showing the ligature c̮i (adapted from Tekin 2003: 22, 23). See descriptions in fig. 2, fig. 2.

ORHON TÜRKÇESİ GRAMERİ   37

2. /Aş/ hece işareti Hemçik-Çırgak yazıtında /äş/ ses öbeğini yazmak için de kullanılmıştır:

/Aşntm/ *äş(i)nt(i)m* "(atlarımı) hızla sürdüm" **(Hem.-Çır. 4)**

/baş/ Hece işareti

32. /baş/ hece işareti ya da ideogramı Tunyukuk **yazıtında bir kez,** bazı Yenisey yazıtlarında ise birkaç kez geçer.

/IBRbaş/ *ı bar baş* "ormanlı doruk" **(T 26)**

/YşKbaş/ *Yaş Ak Baş* (Uyuk-Arcan 5)

/KbaştlıK/ *Ak Baş Atık* (Tuva I 2), vb.

/çi/ Birleşik Harfi

33. /çi/ birleşik harfi Tunyukuk yazıtında bir kez kullanılmıştır:

/kltçimz/ *kältäçim(i)z* (T 14)

/ot/ Hece İşareti

34. /ot/ hece işareti Irk Bitig'de *ot* sözcüğünü yazmak için üç kez kullanılmıştır:

/YSot/ *yaş ot* (IB 17, 53)

/otSUz/ *otsuz* (IB 45)

/Up/ Hece İşareti

35. Hece işareti /Up/ daha çok Irk Bitig'de, seyrek olarak da bazı Yenisey yazıtlarında ve elyazmalarında kullanılmıştır. Bu işaret /up/ ve /üp/ ses öbekleri ile /u/ ünlüsünden sonra gelen /p/ ünsüzünü yazmak için kullanılır.

1. /up/ ses öbeği değerinde:

/ULRUpN/ *ol(o)rupan* "tahta oturup" (IB 28)

/BULUp/ *bolup* "olup" (Elegest I 8), vb.

2. /üp/ ses öbeği değerinde:

/yÜtÜrUp/ *yütürüp* "yitirip" (IB 24)

3. /u/ ünlüsünden sonra gelen /p/ ünsüzü değerinde:

/KUNUUpN/ *konupan* "konup" (IB 64)

Figure 5: Description of the ligature *ci* (from Tekin 2003: 37).