Proposal to encode the Old Turkic ligature ORKHON CI

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

February 17, 2019

1 Proposed character

<table>
<thead>
<tr>
<th>Glyph</th>
<th>Codepoint</th>
<th>Character name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ɽ</td>
<td>10C49</td>
<td>OLD TURKIC LIGATURE ORKHON CI</td>
</tr>
</tbody>
</table>

2 Description

The character Ɽ represents the syllable [či]. In this proposal, it is transliterated as ci, where the tie indicates that the value is represented by a single character. The Ɽ 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:

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

yöyän bolsär : üzugüük : ^=r mülṣ : öğrä : qiiūnda : b̥r̥jā : t̥b̥y̥č̥da : qur̥i̥ja : qur̥d̥nta : jiī̥ra : oγ̥zda : ḷi üč biŋ : sümëz : k̥l̥t̥c̥imëz : b̥r̥ mu nā : ㏁c̥a ㏄b̥nt̥m :
The character \( \mathcal{L} \) is a ligature of \( \U+10C32 \) OLD TURKIC LETTER ORKHON EC + \( \U+10C03 \) OLD TURKIC LETTER ORKHON I, in which \( \mathcal{L} \) is incorporated into the vertical stroke of \( \mathcal{L} \). The \( \mathcal{L} \) occurs simultaneously with \( \mathcal{L} \), which is used throughout the inscription for the normative representation of \( [či] \). The occurrence of both \( \mathcal{L} \) and \( \mathcal{L} \) in the Tonyukuk inscription (lines 11–14, 1b south) is shown below:

The conventional representation of \( [či] \) occurs in:

\[
\begin{align*}
\mathcal{L} & \mathcal{L} \mathcal{T} \mathcal{R} \mathcal{n} \mathcal{G} \mathcal{L} \:
\end{align*}
\]

(line 11) and:

\[
\begin{align*}
\mathcal{L} & \mathcal{L} \mathcal{M} \mathcal{N} \:
\end{align*}
\]

(line 13), and the ligature in:

\[
\begin{align*}
\mathcal{L} \mathcal{L} \mathcal{C} \mathcal{L} \:
\end{align*}
\]

(line 13), and the ligature in:

\[
\begin{align*}
\mathcal{L} \mathcal{L} \mathcal{C} \mathcal{L} \:
\end{align*}
\]

while \( \mathcal{L} \) has the same value as the sequence \( \mathcal{L} \), it is graphically distinct from the latter.

The \( \mathcal{L} \) is similar in appearance to \( \U+10C10 \) 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 \( \mathcal{L} \), the right arm connects in the shape of the head. In \( \mathcal{L} \), the right arm connects at the top of the descender, as in \( \mathcal{L} \) ORKHON I, while in \( \mathcal{L} \) the right arm connects to another appendage that extends from a shortened descender. In any case, the morphological contexts of \( \mathcal{L} \) indicate that it is certainly a ligature of \( \mathcal{L} + \mathcal{L} \), and not the usage of \( \mathcal{L} \) as a proxy for \( [či] \).

The \( \mathcal{L} \) is a graphemic \textit{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 \( \mathcal{L} + \mathcal{L} = \mathcal{L} \). 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 \( \mathcal{L} \), 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 \( \mathcal{L} \mathcal{L} \mathcal{T} \mathcal{R} \mathcal{n} \mathcal{G} \mathcal{L} \), but went from etching \( \mathcal{L} \) to \( \mathcal{L} \), forgetting the intervening \( \mathcal{L} \) (eg. \( \mathcal{L} \mathcal{L} \mathcal{L} \mathcal{T} \mathcal{R} \mathcal{n} \mathcal{G} \mathcal{L} \)), and saw an opportunity in the complementary descenders of the two letters to fix the error using \( \mathcal{L} \mathcal{T} \mathcal{R} \mathcal{n} \mathcal{G} \mathcal{L} \). Whatever the reason may be, the ligation of these two letters into \( \mathcal{L} \) signifies a scribal intentionality that departs from the orthographic conventions of the script, but one that is nonetheless unique and distinctive.

The \( \mathcal{L} \) 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 \( [či] \) (\( \mathcal{L} \) and \( \mathcal{L} \)). Therefore, it is necessary to encode \( \mathcal{L} \) as a new character in the ‘Old Turkic’ block in Unicode.

3 Character Data

Character Properties In the format of UnicodeData.txt:

\[
\begin{align*}
10C49;OLD TURKIC LIGATURE ORKHON CI;Lo;0;R;;;;;;;;;
\end{align*}
\]

Linebreaking In the format of LineBreak.txt:

\[
\begin{align*}
10C49;AL \# Lo \:
\end{align*}
\]
4 References


5 Acknowledgments

I thank Dr. Deborah Anderson of the Script Encoding Initiative (SEI) for providing me with scholarly materials on Old Turkic.
Figure 1: Reproduction of the southern face of Pillar 1 of the Tonyukuk inscription (from Sprengling 1939). The arrow shows the ligature \( c \) 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.\textsuperscript{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 $\text{lt}$, $\text{nt}$, $\text{oq/uq/øq/}\text{qu}$, $\text{šk/šk/šk/šk}$, $\text{šq/qš}$, $\text{iš}$, $\text{aš}$, $\text{baš}$, $\text{up}$, and $\text{ot}$ cannot be regarded as ligatures, because none of them seems to be a combination of two letters.\textsuperscript{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 $\text{š}$ (back $\gamma$), $\text{oq/uq}$ and $\text{ba}$ might have been ideographic in character, e.g., $\text{av}$ = moon, $\text{oq}$ = arrow and $\text{ba}$ = house.\textsuperscript{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., $\text{beth}$ means $\text{ba}$, $\text{bi}$, $\text{bu}$, etc., while in the Turkic

\textsuperscript{11} Thomsen, Inscr., pp. 50, 51; G. Clauson, op. cit., p. 74.

\textsuperscript{12} The only ligature used in the Orkhon inscriptions is $\text{ṭ}=\text{ṭi}$ (see 1.2224). This ligature is obviously a combination of the letters $\text{ṭ}=\text{ṭi}$ and $\text{i}=\text{i}$.

\textsuperscript{13} Thomsen, L’Alphabet runiforme turc, in: Afh., p. 78.

Figure 2: Note regarding $\text{ci}$ as the only attested ligature in the Orkhon inscriptions (from Tekin 1968: 28).
Proposal to encode the Old Turkic ligature ORKHON CI

Anshuman Pandey

1. 22234. Use of the Syllabic Sign ıc

The syllabic sign ıc is rarely used to designate the initial sound group ıc: ıcgrɛt m ıcgɛrtim (BK E 25); ıcɛkd i ıçikdi (BK E 37; TI N 4), ıçkmış ıçikmiš (BK E 9), but generally ıčk- ıčk-; ıča ıčra (TI N 10), but generally ıča.

1. 22235. Use of the Syllabic Sign ıa

The syllabic sign ıa is used once in the inscription of Tonyukuk: ıșg ıșç (TI S 1).

1. 22236. Use of the Syllabic Sign ıa

The syllabic sign ıa is used only once in the inscription of Tonyukuk: ıa (TI N 2).

1. 22237. Use of the Ligature ı (ıç)

In the following word the medial sound group ıç is designated with the ligature ı: ıltɛtɛrmz kältɛçimiz (TI S 7).

1. 2224. Use of Compound Consonant-Characters

1. 22241. Use of the Sign ı

The compound consonant-character ı is used to represent the sound groups alt and ıt in back-vocalic words:

1°. alt: ılm alɛt (OR 1), but generally ılt m, ıltɛm, ıltɛ; ıltɛd ısdı (KT E 38); ıltɛ çaltaçı (KT N 9; TI S 6), but ıltɛt m qalɛt (BK E 14);

2°. ıt: ıltɛ çoltaçı (generally), but ıltɛt (KT N 11; BK E 31); ıltɛ boltsı (generally), but ıltɛt (ıtl) (KC W 3), ıltɛt ıslıta (OF 1, 6), ıltɛt ısinja (KT N 45).

Figure 3: Description of the usage and occurrence of the ligature ıç (from Tekin 1968: 45).
### Proposal to encode the Old Turkic ligature ORKHON CI

Anshuman Pandey

**Figure 4:** Chart of Orkhon letters showing the ligature `cığ` (adapted from Tekin 2003: 22, 23). See descriptions in fig. 2, fig. 2.
Proposal to encode the Old Turkic ligature ORKHON CI

Anshuman Pandey

Proposta de codificar a ligatura antiga turca ORKHON CI

Anshuman Pandey

2. /AŞ/ hece işaretleri Hemçik-Çırgak yazıtında /AŞ/ ses öbeğini yazmak için de kullanılmıştır:
   /AŞNTM/ "atlanımlı" hızla sürdüm" (Hem.-Çır. 4)

/baŞ/ Hece İşareti

32. /baŞ/ hece işaretleri ya da ideogramı Tunyukuk yazıtında bir kez, bazı Yenisey yazıtlarında ise birkaç kez geçer.
   /IBRbaŞ/ bar baŞ “ormanlı doruk” (T 26)
   /YŞKbaŞ/ Yaş Ak BaŞ (Uyuk-Arcan 5)
   /KbaşlıK/ Ak BaŞ Atık (Tuva I 2), vb.

/ci/ Birleşik Harfı

33. /ci/ birleşik harf Tunyukuk yazıtında bir kez kullanılmıştır:
   /kltčim/ kältčim(i)z (T 14)

/ot/ Hece İşareti

34. /ot/ hece işaretleri Irk Bitig’de ot sözçüğünü yazmak için üç kez kullanılmıştır:
   /YSoT/ yaş ot (IB 17, 53)
   /otSUz/ otsuz (IB 45)

/Up/ Hece İşareti

   1. /up/ ses öbeği degerinde:
      /ULRUUpN/ ol(o)rapan “tahta oturup” (IB 28)
      /BULUp/ bolup “olup” (Elegez I 8), vb.
   2. /üp/ ses öbeği degerinde:
      /yÜTRUp/ yütürüp “yitip” (IB 24)
   3. /u/ ünlüsünden sonra gelen /p/ ünlüsü degerinde:
      /KUNUUpN/ konupan “konup” (IB 64)

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