Unicode request for affricate ligatures

Kirk Miller, kirkmiller, gmail.com 2024 February 21

This proposal, officially supported by the International Phonetic Association after evaluation by the IPA Alphabets, Charts and Fonts Committee (Nicolaides 2024), requests old-style IPA ligatures for the coronal (d- and t-type) affricates that are not supported by Unicode.

Before the 1989 Kiel Convention, the IPA had an established pattern of ligatures to transcribe affricates (sounds such as English ch [ʧ] and j [ʤ]). The series was open-ended: the 1949 Principles said only (p. 14–15),

>Digraphs] To represent affricates, e.g. pf, bv, ts, dz, tf̂, dʒ̂, te, kx. If a language contains affricates as well as such sequences as t + s, t + f̂, the affricates may be denoted by ligature forms such as ts, dz, tf̂, dʒ̂, or by the use of a linking mark ̂ or ...

Similarly, the 1978 chart (JIPA, vol. 8, no. 1/2) said,

AFFRICATES can be written as digraphs, as ligatures, or with slur marks; thus ts, tf̂, dʒ̂: ts, tf̂, dʒ̂.

These conventions were illustrated with the most common affricates, which were sibilants, but all affricates were implied, and non-sibilant ligatures are found in the literature. Unicode currently only supports ligatures for sibilant affricates, namely ⟨ʦ ʣ, ʧ ʤ, ꭧ ꭦ, ʨ ʥ⟩.

In 1989, the IPA officially retired ligatures, leaving only the tie-bar to formally distinguish affricates from sequences. The ligature ⟨ʧ⟩, for example, was replaced with the digraph ⟨tʃ⟩ (or equivalently ⟨tʃ⟩). However, old-style ligatures such as ⟨ʧ⟩ continue to be used by linguists who do not care for the aesthetics of the tie-bar, which can clutter up transcription. It is quite common for the tie-bar to simply be omitted, per the 1949 Principles, but this can be a problem for a language like Polish that contrasts affricates such as /tʃ/ ~ /ʧ/ (a single consonant) and sequences such as /tʃ/ (two consonants).

Among non-sibilant affricates, the voiceless alveolar lateral [ɗ] is quite common in the world’s languages, though its voiced homologue [ʤ] is relatively rare. Nonetheless, we can attest to both of these ligatures in Sandawe /ɗ/ and /ʤ/ (see figures).

Voiced and voiceless dental affricates [ɓ ɗ] are found i.a. in Burmese and Western Nilotic languages. [ʈʰ] occurs as well in many northern Athabaskan languages, and the aspiration distinction of Athabaskan /ʈʰ/ is frequently transcribed as one of voicing, /ɓ ɗ/. We have not
yet attested to these as ligatures in the literature; less precise ⟨dʰ tʰ⟩ or older IPA ⟨d⁵ t⁵⟩ tend to be used instead.

The retroflex laterals [ʈʃ] and [ɖʐ], on the other hand, are quite rare. They occur phonemically in the Pahari language Bhadrawahi for old *Cr clusters and phonetically in the Nuristani language Kāmviri as the realization of the underlying clusters /ʈl/, /ɖɭ/. Richard Strand, an expert on Nuristani languages, said (p.c., 2021 Jan 26) that he would be glad to use the ligatures:

I’m for getting rid of the tie bars for all affricates … ligatures for all the apico-postalveolar affricates ([ʈʃ] [ɖʐ] [ʈʂ] [ɖʐ]) would be welcomed.

(Ligatures for the first pair, [ʈʃ] [ɖʐ], have since been added to Unicode at U+AB66 ʈʂ and AB67 ɖʐ. Note that only a single retroflex hook is needed, because by the definition that an affricate has a single place of articulation, it applies to the ligature as a whole.)

These six ligatures are the only possibilities among the pulmonic coronal consonants of the current IPA alphabet. Extension to ejective affricates such as ⟨ʈʂ’⟩ and ⟨ɖʐ’⟩ is trivial, and implosive affricates do not occur.

Ligatures for non-coronal affricates, such as *⟨ʃf ḥ, ɬv, αẓ jı, ˚c, kx qı, ks, qı ϛ⟩, or for doubly articulated consonants such as *⟨kpf ɬf ɹ⟩, are not attested from the literature. Among historical IPA letters, only ⟨ʈf ɖg⟩ are requested (see next paragraph) – historical *⟨ʈs ɖʃ⟩ are unattested, and the fricative letters ⟨ʃ ɹ⟩, once used for Shona, are obsolescent.

In IPA publications from 1947 to 1989, there are “special forms” of palatalized ⟨ʃ⟩ and ⟨ʒ⟩ that have a curl rather than a hook, namely U+0286 ʃ and U+0293 ʒ for hooked U+1D8B ʃ and U+1DF18 ʒ. The corresponding affricate ligatures ⟨ʈʃ⟩ (which is found in the IPA journal) and ⟨ɖg⟩ are therefore proposed here as characters distinct from U+1DF17 ʈʃ and U+1DF12 ɖg. Although only the voiceless member of the pair is attested, due to the relative infrequency of [ɖg] in Russian and Ukrainian (the languages that [ʈʃ] is attested for), both characters are requested because support for one alone would limit its utility. Besides the official support of the IPA, see the argument made in L2/20-004 Unicode request for dezh with retroflex hook for the identical situation with U+1DF19 ɖg.

We therefore request Unicode support for the two historical sibilant ligatures ⟨ɖg ʃ⟩ and the six modern non-sibilant coronal ligatures ⟨ʈʃ ʈʂ, ɖʐ ɖʐ, ʈʂ’ ɖʂ’⟩.
Characters

The proposed names continue the established IPA misnomer of calling ligatures ‘digraphs.’

Affricate ligatures

- 1DF1F LATIN SMALL LETTER D-ETH DIGRAPH.
- 1DF20 LATIN SMALL LETTER D-LEZH DIGRAPH.
- 1DF21 LATIN SMALL LETTER D-LEZH DIGRAPH WITH RETROFLEX HOOK.
- 1DF22 LATIN SMALL LETTER TL DIGRAPH WITH BELT.
- 1DF23 LATIN SMALL LETTER TL DIGRAPH WITH RETROFLEX HOOK AND BELT.
- 1DF24 LATIN SMALL LETTER T-THEETA DIGRAPH.

Ligatures of historical IPA letters

- 1DF2B LATIN SMALL LETTER DEZH DIGRAPH WITH CURL.
- 1DF2C LATIN SMALL LETTER TESH DIGRAPH WITH CURL.

DoNotEmit data

For historical reasons, IPA letters with retroflex hook are not canonically equivalent to the letter plus the retroflex hook diacritic. They should thus be listed in DoNotEmit.txt.

1DF20 0322; 1DF21; Precomposed_Form # LATIN SMALL LETTER D-LEZH DIGRAPH, COMBINING RETROFLEX HOOK BELOW; LATIN SMALL LETTER D-LEZH DIGRAPH WITH RETROFLEX HOOK
1DF22 0322; 1DF23; Precomposed_Form # LATIN SMALL LETTER TL DIGRAPH WITH BELT, COMBINING RETROFLEX HOOK BELOW; LATIN SMALL LETTER TL DIGRAPH WITH RETROFLEX HOOK AND BELT

Properties

1DF1F;LATIN SMALL LETTER D-ETH DIGRAPH;Ll;0;L;;;;N;;;;;
1DF20;LATIN SMALL LETTER D-LEZH DIGRAPH;Ll;0;L;;;;N;;;;;
1DF21;LATIN SMALL LETTER D-LEZH DIGRAPH WITH RETROFLEX HOOK;Ll;0;L;;;;N;;;;;
1DF22;LATIN SMALL LETTER TL DIGRAPH WITH BELT;Ll;0;L;;;;N;;;;;
1DF23;LATIN SMALL LETTER TL DIGRAPH WITH RETROFLEX HOOK AND BELT;Ll;0;L;;;;N;;;;;
1DF24;LATIN SMALL LETTER T-THEETA DIGRAPH;Ll;0;L;;;;N;;;;;
1DF2B;LATIN SMALL LETTER DEZH DIGRAPH WITH CURL;Ll;0;L;;;;N;;;;;
1DF2C;LATIN SMALL LETTER TESH DIGRAPH WITH CURL;Ll;0;L;;;;N;;;;;
References


**Chart**

Greyed out cells are already assigned.

<table>
<thead>
<tr>
<th>Latin Extended-G</th>
<th>1DF00</th>
<th>1DF1</th>
<th>1DF2</th>
<th>...</th>
<th>1DFF</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1Fj</td>
<td>x</td>
<td>1Fj</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>e</td>
<td>h</td>
<td>1Fg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>e</td>
<td>1Fg</td>
<td>t</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>x</td>
<td>1Fg</td>
<td>t</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>e</td>
<td>1Fg</td>
<td>t(\theta)</td>
<td>      </td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1Fg</td>
<td>1Fg</td>
<td>r</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>x</td>
<td>1Fg</td>
<td>l</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>e</td>
<td>1Fg</td>
<td>n</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1Fg</td>
<td>r</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>e</td>
<td>1Fg</td>
<td>s</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>1</td>
<td>i</td>
<td>t</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>e</td>
<td>o</td>
<td>1Fg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>e</td>
<td>1Fg</td>
<td>1Fg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>c</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td>s</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>c</td>
<td>d</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figures

<table>
<thead>
<tr>
<th>fricative</th>
<th>f</th>
<th>s</th>
<th>ɬ</th>
<th>x</th>
<th>h</th>
</tr>
</thead>
<tbody>
<tr>
<td>glottalised</td>
<td>ɽ̅</td>
<td>ʃ̅</td>
<td>ʃ</td>
<td>ʰ</td>
<td>k̂</td>
</tr>
<tr>
<td>aspirated</td>
<td>ɽ̅</td>
<td>ʃ̅</td>
<td>ɬ</td>
<td>ʰ</td>
<td>k̂</td>
</tr>
<tr>
<td>voiceless</td>
<td>ɽ</td>
<td>ʃ</td>
<td>ɬ</td>
<td>ʰ</td>
<td>k̂</td>
</tr>
<tr>
<td>voiced</td>
<td>ɽ</td>
<td>ʃ</td>
<td>ɬ</td>
<td>ʰ</td>
<td>k̂</td>
</tr>
<tr>
<td>nasal</td>
<td>ɽ</td>
<td>ʃ</td>
<td>ɬ</td>
<td>m</td>
<td>n</td>
</tr>
</tbody>
</table>

Figure 1. Elderkin (1989: 37). The tenuis (‘voiceless’) consonants are here written with a redundant voiceless ring diacritic, so that a forgotten aspiration or voicing diacritic will not result in a misreading. The tenuis lateral affricate is written as a t-ɬ ligature rather than as a t-ɬ ligature like the glottalized affricate. Since ⟨t̅l⟩ would be imprecise notation for a true affricate, we do not request a t-ɬ ligature. (Elderkin may be indicating that /t̅l/ does not have the fully fricated release of /t̅l̆/.)

Figure 2. Elderkin (1989: 23). List of IPA ligatures for Sandawe.

Figure 3. Elderkin (1982: 7).

Figure 4. Elderkin (1989: 119).

Figure 5. Elderkin (1989: 2).
phonetically more like [tʂ]. The articulation of 𓊁 has noticeably more friction than hes ǂ which rarely does have any, but all of the three stops in the central series have equal amounts of friction. Copland 1938 notes lack of affrication equally with 𓊁 as well as

Figure 6. Elderkin (1989: 36).

<table>
<thead>
<tr>
<th>𓊁</th>
</tr>
</thead>
<tbody>
<tr>
<td>𓊁ànì arrow</td>
</tr>
<tr>
<td>𓊁áxáxá kichomi</td>
</tr>
<tr>
<td>𓊁ãːts'âːtë tree sp used for mishale</td>
</tr>
<tr>
<td>𓊁ðômóː 𓊁ðôméː kununua</td>
</tr>
</tbody>
</table>

Figure 7. Elderkin (ms, no date). Beginning of the ⟨dh⟩ entry of a ca. 2300-item Sandawe wordlist.

<table>
<thead>
<tr>
<th>ǂ</th>
</tr>
</thead>
<tbody>
<tr>
<td>ǂàbísó intestines, stomach</td>
</tr>
<tr>
<td>ǂãːgáː tree sp</td>
</tr>
<tr>
<td>ǂãːkímé shoe</td>
</tr>
</tbody>
</table>

Figure 8. Ibid. Beginning of the ⟨d⟩ entry of the Sandawe wordlist.

<table>
<thead>
<tr>
<th>ǂːːći (perf) kuroga; ǂːːwàːs roga many people</th>
</tr>
</thead>
<tbody>
<tr>
<td>ǂːːsë mroga; ǂːːʔːsé niroga;</td>
</tr>
</tbody>
</table>

Figure 9. Ibid. Parallel ligatures ⟨t⟩ and ⟨ts’⟩.
Figure 10. Elderkin (1978: 5, 18). Hadza does not have the voiced affricate, but here the ⟨ṭ⟩ ligature is used for both /d'/ and /d/ alongside /ts ts ʧ ʤ/.  

In fact two words show this collocation, ḥanu 'ponis' and ḏa'l 'nyrax'. The first is one of the three  

Figure 11. Elderkin (1988: 79). ⟨ṭ⟩ for Iraqw.  

Figure 12. de Bray (1951: 75). Curly-tail ⟨ʧ⟩ for Russian.  

Figure 13. Partridge (1951: 47). ⟨ʧ⟩ in IPA transcription of Pushkin.
**A. Administrative**

<table>
<thead>
<tr>
<th>1. Title:</th>
<th>Affricate ligatures</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Requester's name:</td>
<td>Kirk Miller</td>
</tr>
<tr>
<td>3. Requester type (Member body/Liaison/Individual contribution):</td>
<td>individual</td>
</tr>
<tr>
<td>4. Submission date:</td>
<td>2024 February 21</td>
</tr>
<tr>
<td>5. Requester's reference (if applicable):</td>
<td></td>
</tr>
</tbody>
</table>

Choose one of the following:

| (or) More information will be provided later: | yes |

**B. Technical – General**

1. Choose one of the following:

   - a. This proposal is for a new script (set of characters):
     - Proposed name of script: |
   - b. The proposal is for addition of character(s) to an existing block:
     - Name of the existing block: Latin Extended-G |

2. Number of characters in proposal: 8

3. Proposed category (select one from below - see section 2.2 of P&P document):

   - A-Contemporary x
   - B.1-Specialized (small collection) |
   - B.2-Specialized (large collection) |
   - C-Major extinct |
   - D-Attested extinct |
   - E-Minor extinct |
   - F-Archaic Hieroglyphic or Ideographic |
   - G-Obscure or questionable usage symbols |

4. Is a repertoire including character names provided?

   | if YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document? | yes |

5. Fonts related:

   a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?
      | Kirk Miller |

   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
      | SIL (Gentium Release) |

6. References:

   a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? yes

   b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? yes

7. Special encoding issues:

   Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)? yes

8. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at www.unicode.org for such information on other scripts. Also see Unicode Character Database (www.unicode.org/reports/tr44/) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

---

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?
   - Yes: no

2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?
   - Yes: yes
   - If yes, available relevant documents: [see letter of support]
   - If yes, with whom: The International Phonetic Organization

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?
   - Reference: 

4. The context of use for the proposed characters (type of use; common or rare)
   - Reference: phonetic

5. Are the proposed characters in current use by the user community?
   - Yes
   - Reference: see illustrations

6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?
   - Yes
   - If yes, is a rationale provided?

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?
   - Yes
   - If yes, is a rationale for its inclusion provided?

8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?
   - Yes
   - If yes, is a rationale for its inclusion provided?

9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?
   - Yes
   - If yes, is a rationale for its inclusion provided?

10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character?
    - Yes
    - If yes, is a rationale for its inclusion provided?

11. Does the proposal include use of combining characters and/or use of composite sequences?
    - Yes
    - If yes, is a rationale for such use provided?
    - Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

12. Does the proposal contain characters with any special properties such as control function or similar semantics?
    - Yes
    - If yes, describe in detail (include attachment if necessary)

13. Does the proposal contain any Ideographic compatibility characters?
    - Yes
    - If yes, are the equivalent corresponding unified ideographic characters identified?