

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 (Nicolaidis 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* [tʃ] and *j* [dʒ]). The series was open-ended: the 1949 *Principles* said only (p. 14–15),

[Digraphs] To represent affricates, e.g. **pf, bv, ts, dz, tʃ, dʒ, tʂ, kx**. If a language contains affricates as well as such sequences as **t + s, t + ʃ**, the affricates may be denoted by ligature forms such as **ts, dz, tʃ, dʒ**, or by the use of a linking mark $_$ or $\widehat{\quad}$...

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, tʃ, dʒ: ts, tʃ, dʒ: \widehat{ts} , $\widehat{tʃ}$, $\widehat{dʒ}$.

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 $\langle ts dz, tʃ dʒ, tʃ dʒ, tʃ dʒ, tʃ dʒ \rangle$.

In 1989, the IPA officially retired ligatures, leaving only the tie-bar to formally distinguish affricates from sequences. The ligature $\langle tʃ \rangle$, for example, was replaced with the digraph $\langle tʃ \rangle$ (or equivalently $\langle \widehat{tʃ} \rangle$). However, old-style ligatures such as $\langle tʃ \rangle$ 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/ ~ /tʃ/ (a single consonant) and sequences such as /tʃ/ (two consonants).

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

Voiced and voiceless dental affricates [dʒ tʃ] are found *i.a.* in Burmese and Western Nilotic languages. [tʃ] occurs as well in many northern Athabaskan languages, and the aspiration distinction of Athabaskan /tʃ tʃʰ/ is frequently transcribed as one of voicing, /dʒ tʃ/ . We have not

yet attested to these as ligatures in the literature; less precise <d^δ t^θ> or older IPA <d^δ t^θ> tend to be used instead.

The retroflex laterals [ɬ̺] and [ɮ̺], 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 Kām̐viri as the realization of the underlying clusters /t̺l/, /d̺l/. 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 ([t̺ʂ] [d̺ʂ] [t̺ʈ] [d̺ʈ]) would be welcomed.

(Ligatures for the first pair, [t̺ʂ] [d̺ʂ], have since been added to Unicode at U+AB66 t̺ʂ and AB67 d̺ʂ. 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 <t̺ʰ'> and <t̺ʰ'> is trivial, and implosive affricates do not occur.

Ligatures for non-coronal affricates, such as *<ɸ̠ ɸ̠, ɸ̠ ɸ̠, ɸ̠ ɸ̠, ɸ̠ ɸ̠, ɸ̠ ɸ̠, ɸ̠ ɸ̠>, or for doubly articulated consonants such as *<ɸ̠ ɸ̠ ɸ̠>, are not attested from the literature. Among historical IPA letters, only <ɸ̠ ɸ̠> are requested (see next paragraph) – historical *<t̺ʂ ɸ̠> are unattested, and the fricative letters <ʂ ʂ>, 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 <ʃ̣> and U+0293 <ʒ̣> for hooked U+1D8B <ʃ̣> and U+1DF18 <ʒ̣>. The corresponding affricate ligatures <ɸ̠> (which is found in the IPA journal) and <ɸ̠> are therefore proposed here as characters distinct from U+1DF17 <ɸ̠> and U+1DF12 <ɸ̠>. Although only the voiceless member of the pair is attested, due to the relative infrequency of [ɸ̠] in Russian and Ukrainian (the languages that [ɸ̠] 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 deẓh with retroflex hook* for the identical situation with U+1DF19 <ɸ̠>.

We therefore request Unicode support for the two historical sibilant ligatures <ɸ̠ ɸ̠> and the six modern non-sibilant coronal ligatures <t̺ ɸ̠, ɸ̠ ɸ̠, ɸ̠ ɸ̠, ɸ̠ ɸ̠, ɸ̠ ɸ̠, ɸ̠ ɸ̠>.

Characters

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

Affricate ligatures

- ɖ 1DF1F LATIN SMALL LETTER D-ETH DIGRAPH.
- ɗ 1DF20 LATIN SMALL LETTER D-LEZH DIGRAPH.
- ɗ̥ 1DF21 LATIN SMALL LETTER D-LEZH DIGRAPH WITH RETROFLEX HOOK.
- ɥ 1DF22 LATIN SMALL LETTER TL DIGRAPH WITH BELT.
- ɥ̥ 1DF23 LATIN SMALL LETTER TL DIGRAPH WITH RETROFLEX HOOK AND BELT.
- θ 1DF24 LATIN SMALL LETTER T-THETA DIGRAPH.

Ligatures of historical IPA letters

- ɖ͡ 1DF2B LATIN SMALL LETTER DEZH DIGRAPH WITH CURL.
- ɥ͡ 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-THETA 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

E.D. Elderkin (1978) *Loans in Hadza*. University of Dar es Salaam.

———— (1982) *Some Preliminary Notes on Sandawe Tone*. University of Dar es Salaam.

———— (1988) *Person and Number Markers in Iraqw Verbs*. AAP 14: 79–96.

———— (1989) *The Significance and Origin of the Use of Pitch in Sandawe*. University of York.

Katerina Nicolaidis (2024) ‘Unicode support for historical and para-IPA letters.’ Letter submitted to the Unicode Technical Committee, 01 January 2024. [L2/24-049](#).

Richard Strand (2010) ‘Nurestâni Languages,’ in *Encyclopaedia Iranica*.

Chart

Greyed out cells are already assigned.

Latin Extended-G

1DF00 1DFFF

	1DF0	1DF1	1DF2	...	1DFF
0	ƒ	ƒ	ƒ		
1	ƒ	ƒ	ƒ		
2	ƒ	ƒ	ƒ		
3	ƒ	ƒ	ƒ		
4	ƒ	ƒ	ƒ		
5	ƒ	ƒ	ƒ		
6	ƒ	ƒ	ƒ		
7	ƒ	ƒ	ƒ		
8	ƒ	ƒ	ƒ		
9	ƒ	ƒ	ƒ		
A	ƒ	ƒ	ƒ		
B	ƒ	ƒ	ƒ		
C	ƒ	ƒ	ƒ		
D	ƒ	ƒ			
E	ƒ	ƒ			
F	ƒ	ƒ			

Figures

fricative				f	s	ʃ	x	h
glottalised	çʷ	ʃʷ	ʒʷ		ʃʷ	ʃʷ	kʷ	ʔ
aspirated	çʰ	ʃʰ	ʒʰ	pʰ	tʰ	ʃʰ	kʰ	
voiceless	ç	ʃ	ʒ	p	t	ʃ	ʈ	k
voiced	ç	ʃ	ʒ	b	d	ʒ	dʒ	g
nasal	ɕ	ʒ	ʒ	m	n			

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-l ligature rather than as a t-ʃ ligature like the glottalized affricate. Since <ʈ> would be imprecise notation for a true affricate, we do not request a t-l ligature. (Elderkin may be indicating that /ʈ/ does not have the fully fricated release of /ʃʷ/.)

Ligatures: ʃ **ʃʷ** ʈ **dʒ**

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

ʃak'ime ʒa:ʔwasi ʔiʔwa:
 shoe good pl 1Sg 3m
 Sg pl perf
 'I gave him some good shoes'

Figure 3. Elderkin (1982: 7).

ʔàtá 'dìʔáʔ. **'dʒ**òmó
 father really knows how to buy eggs

Figure 4. Elderkin (1989: 119).

'híáʔ. **²ʃ**èsé: 'wagògá: ²héʔékàjʔ. 'tétà 'lõ:kó 'tétà
 ²fákáʔ.wà:j

Figure 5. Elderkin (1989: 2).

phonetically more like [tʒ]. The articulation of **ɖʒ** has noticeably more friction than has ɖ 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 **ɖʒ** as well as

Figure 6. Elderkin (1989: 36).

ɖʒ
 ɖʒàní arrow
 ɖʒáxáxá kichomi
 ɖʒǎ:t'àt'é tree sp used for mishale
 ɖʒòmó: ɖʒòmé: kununua

Figure 7. Elderkin (ms, no date). Beginning of the <ɖʒ> entry of a ca. 2300-item Sandawe wordlist.

ɖ'
 ɖ'ábísó intestines, stomach
 ɖ'ágá: tree sp
 ɖ'àkìmé shoe

Figure 8. *Ibid.* Beginning of the <ɖ'> entry of the Sandawe wordlist.

ɖ'é:tsì (perf) kuroga; ɖ'é:wàts' roga many people
 ɖ'é:tsě mroga; ɖ'é:ʔ'sé niroga;

Figure 9. *Ibid.* Parallel ligatures <ɖ'> and <ts'>.

Table I
The nonclick consonants
of Hadzo

	ʋ	ʈ	ʃ	k'	k''
f	s	ʈ	ʃ		h
p	t	ʈ	ʃ	ɬ	k' ?
b	d	ɖ	ɖ	ɖ	ɖ

In fact two words show this collocation, **ʈuʈu** 'penis' and **ʈaʈa** 'nyrax'. The first is one of the three

Figure 10. Elderkin (1978: 5, 18). Hadza does not have the voiced affricate, but here the <ʈ> ligature is used for both /ʈ'/ and /ʈ/ alongside /ts' ts dz tʃ tʃ dz/.

<p><i>Consonants</i></p> <table style="margin-left: 20px;"> <tr> <td rowspan="3" style="font-size: 2em; vertical-align: middle;">{</td> <td style="padding-right: 10px;">ejective</td> <td style="padding-right: 10px;">ʈ</td> <td style="border: 1px solid red; padding: 2px;">ʈ</td> <td style="padding-right: 10px;">q</td> <td style="padding-right: 10px;">q̄</td> </tr> <tr> <td>plosive</td> <td>p</td> <td>t</td> <td>k</td> <td>k̄</td> </tr> <tr> <td>fricative</td> <td>f</td> <td>s</td> <td>ʃ</td> <td>x</td> </tr> </table>	{	ejective	ʈ	ʈ	q	q̄	plosive	p	t	k	k̄	fricative	f	s	ʃ	x	<table style="border: 1px solid red; padding: 5px;"> <tr><td>1Sg</td><td>ʈáw</td></tr> <tr><td>2Sg</td><td>ʈé:r</td></tr> <tr><td>3Sg</td><td>ʈáj</td></tr> <tr><td>1Pl</td><td>ʈàwá:n</td></tr> <tr><td>2Pl</td><td>ʈè:rá?</td></tr> <tr><td>3Pl</td><td>ʈájá?</td></tr> <tr><td></td><td>go</td></tr> </table>	1Sg	ʈáw	2Sg	ʈé:r	3Sg	ʈáj	1Pl	ʈàwá:n	2Pl	ʈè:rá?	3Pl	ʈájá?		go
{		ejective	ʈ	ʈ	q	q̄																									
		plosive	p	t	k	k̄																									
	fricative	f	s	ʃ	x																										
1Sg	ʈáw																														
2Sg	ʈé:r																														
3Sg	ʈáj																														
1Pl	ʈàwá:n																														
2Pl	ʈè:rá?																														
3Pl	ʈájá?																														
	go																														
<p>the alveolar is affricated, the velar an no ʈ (nor [ɬ]), for the ʈ' to contrast</p>																															

Figure 11. Elderkin (1988: 79). <ʈ> for Iraqw.

All the *chuintantes*, including ч and щ, in contrast to Russian, are hard in Ukrainian, hence чай (= tea) sounds more like Serbocroatian tʃaj (чaj) than Russian **tʃæj** (чай). A slightly softer version is heard before i, as in жінка, (= woman), шість (= six).

Figure 12. de Bray (1951: 75). Curly-tail <tʃ> for Russian.

ra'diʈɪlɪm. a'na **pra'tʃaləs** s 'nɪmɪ f 'samɪx 'troɡətɪlɪtɪx vɪra'zɛnɪjɪx, ɪzɪt'ɪalə svot pra'stupək ɲədə'ɫiməjə 'ʃiləjə 'straʃtɪ ɪ **a'kanʈɪvələ** 'tɛm, ʃtə bla'zɛnɪjʃɪjə mɪ'nutəjə 'zɪzɪ **paʈ'tot** a'na 'tu, kag'da paz'volɪnə 'budɪt 'jej 'broʃɪtsə g na'gam dra'zajʃɪx jɪ'jo ra'diʈɪlɪj. **zəpɪ'tʃatəf** 'obə pʊʃ'ma . . . əna 'broʃɪləs nə paʃ'tɛl pʊʃɪt 'samɪm ras'ʃetəm ɪ zəpɪ'malə.

'pʊʃkɪnə : " 'povɪʃtɪ 'bɛlkɪnə "
 PUSHKIN : " 'teɪlz əv 'bɛlkin."
 transkripʃn bai M. PARTRIDGE.

Figure 13. Partridge (1951: 47). <tʃ> in IPA transcription of Pushkin.

ISO/IEC JTC 1/SC 2/WG 2
**PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
 FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹.**

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html for guidelines and details before filling this form.

Please ensure you are using the latest Form from std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html.
 See also std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html for latest *Roadmaps*.

A. Administrative

1. Title:	<u>Affricate ligatures</u>
2. Requester's name:	<u>Kirk Miller</u>
3. Requester type (Member body/Liaison/Individual contribution):	<u>individual</u>
4. Submission date:	<u>2024 February 21</u>
5. Requester's reference (if applicable):	<u></u>
6. Choose one of the following:	
This is a complete proposal:	<u>yes</u>
(or) More information will be provided later:	<u></u>

B. Technical - General

1. Choose one of the following:	
a. This proposal is for a new script (set of characters):	<u></u>
Proposed name of script:	<u></u>
b. The proposal is for addition of character(s) to an existing block:	<u>yes</u>
Name of the existing block:	<u>Latin Extended-G</u>
2. Number of characters in proposal:	<u>8</u>
3. Proposed category (select one from below - see section 2.2 of P&P document):	
A-Contemporary <input checked="" type="checkbox"/>	B.1-Specialized (small collection) <input type="checkbox"/>
C-Major extinct <input type="checkbox"/>	B.2-Specialized (large collection) <input type="checkbox"/>
D-Attested extinct <input type="checkbox"/>	E-Minor extinct <input type="checkbox"/>
F-Archaic Hieroglyphic or Ideographic <input type="checkbox"/>	G-Obscure or questionable usage symbols <input type="checkbox"/>
4. Is a repertoire including character names provided?	<u>yes</u>
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	<u>yes</u>
b. Are the character shapes attached in a legible form suitable for review?	<u>yes</u>
5. Fonts related:	
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<u>Kirk Miller</u>
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<u>SIL (Gentium Release)</u>
6. References:	
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<u>yes</u>
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<u>yes</u>
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)?	<u>yes</u>

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.

¹ Form number: N4502-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? If YES explain	 	<i>no</i>
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.)? If YES, with whom? If YES, available relevant documents:	 <i>The International Phonetic Organization</i> <i>(see letter of support)</i>	<i>yes</i>
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:	 	<i>phonetic</i>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	 <i>see illustrations</i>	<i>yes</i>
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference:	 	<i>no</i>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?		<i>if possible</i>
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? If YES, reference:	 	<i>no</i>
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference:	 	<i>no</i>
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? If YES, is a rationale for its inclusion provided? If YES, reference:	 	<i>no</i>
11. Does the proposal include use of combining characters and/or use of composite sequences? If YES, is a rationale for such use provided? If YES, reference:	 	<i>no</i>
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? If YES, reference:	 	<i>no</i>
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)	 	<i>no</i>
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:	 	<i>no</i>