

Unicode request for dezh with retroflex hook

Kirk Miller, kirkmiller@gmail.com

Michael Everson, everson@evertype.com

2021 January 03

The original version of L2/20-125r, *Unicode request, Miller 2020, expected IPA retroflex letters and similar hooks*, had requested a pair of letters, <ɖʒ ɥʒ>, that have been used in phonetic descriptions of Polish orthographic *cz* and *dź*. Due to an asymmetry in attestation, only <ɥʒ> was recommended by the SAH for approval by the UTC, which approved it in 2020 July. However, the acceptance of one character without its partner makes it essentially useless, as both are needed if either is to be used for any significant quantity of transcription. John Esling and Michael Ashby of the IPA concur. In its 2020 October meeting, the SAH recommended <ɖʒ> based on the current proposal, and left space for it in a list of characters recommended for the Latin Extended-G block.

Background

Phonetic characters for alveolar stops and affricates come in voicing pairs, reflecting the ubiquity of a voicing distinction in European languages. It would thus hinder the utility of Unicode to encode just one member of a pair for reasons of attestation: each is clearly part of a structured set, and anyone using a transcription system with a voiceless letter for a language like Polish or Russian will need the voiced letter as well. If we do not encode both, the one that has been accepted will not serve any such user.

The form of the letter is self-evident. Without the summary above as a guide, could the UTC tell which of the ligatures below has not yet been attested? (The SAH was not able to do so in September, given that several months had passed since they had last seen the proposal.)

z	s	ʒ	ʂ	ʒ	ʃ
ɖz	ts	ɖʒ	tʂ	ɖʒ	tʃ
z̥	s̥	ʒ̥	ʂ̥	ʒ̥	ʃ̥
ɖz̥	tʃ̥	ɖʒ̥	tʃ̥	tʃ̥	ɖʒ̥

The component characters, U+1D98 \int and 1D9A ʒ , were adopted with proposal L2/03-190r by Peter Constable. Just as here, Constable could only attest to the voiceless consonant (his Figure 7), but the UTC accepted both characters as a pair.

John Esling, former president of the IPA, final editor of the IPA *Handbook* and author of its appendix on the computer coding of IPA symbols, wrote (p.c. 2020 Oct 19),

Laver (1994: 560) gives us (someone's?) transcriptions of a few Polish words. Only the voiceless fricative and affricate in question appear, because the list is short. They both have voiced counterparts, as do all of the (pre-velar) fricatives and affricates in Polish (Jassem, JIPA, 2003). We could say that Polish is a full set of this genre of sibilant contrast, including post-dental (or whatever it is) as a third series. Although the voiced values may not have been put into print in any particular instance, we can assume that Laver would have had occasion to use them had Polish words containing the voiced fricatives and affricates presented themselves in his list.

Now Jassem calls the /tʃ dʒ/ pair alveolar. He calls the /tɕ dʑ/ pair alveolo-palatal. My view is that trying to nail these sounds down to a place of articulation is extremely limiting; the tongue is doing more than that (not to mention the lips). But as I recall, Hamann (JIPA 2003) found evidence of the first pair being retroflex, and Żygis echoes that view: [\(Non\)retroflexivity of Slavic affricates and its motivation: evidence from Polish and Czech <č>](#).

As your comment suggests, not everyone would agree that these sounds are archetypically retroflex (e.g. as in Russian). I like very much that Laver's transcriptions took a 'middle road', between alveolar and retroflex. On this basis, I would support codification of the 'intermediate' symbolic representation (of the 'alveolar' set with the retroflex hook) for the two fricatives and the two affricates in this series (vɭ and vɔ).

That is, I agree with Michael Everson that it only makes sense to include both members of each pair, as it is just as likely (e.g. in Polish) for the vɔ member to occur as the vɭ one, and if one's transcription is so inclined, all four sounds will need symbols to represent them.

Michael Ashby, current president of the IPA, wrote (p.c. 2020 Oct 20),

I take the same view as John and Michael Everson, and I don't mind putting my name to a brief statement if it will help. It only makes sense to include both the voiceless and the voiced together.

Requested character

The referenced codepoint **in red** is newly approved by the SAH but not yet by the UTC.

ɖʒ 1DF19 LATIN SMALL LETTER DEZH WITH RETROFLEX HOOK. See Figure 1.

Forms a pair with **1DF1C** tʃ LATIN SMALL LETTER TESH WITH RETROFLEX HOOK.

Properties

1DF19;LATIN SMALL LETTER DEZH WITH RETROFLEX HOOK;L1;0;L;;;;;N;;;;;

Chart

A gap was left for <ɟ> among the Latin Extended-G codepoints (grey) that were recommended by the SAH on 2020 Oct 19. <ɟ> was not included in Everson’s chart of those characters so as not to interfere with their evaluation by the UTC.

	...0	...1	...2	...3	...4	...5	...6	...7	...8	...9	...A	...B	...C	...D	...E	...F
Latin Extended-G																
U+1DF0x																
U+1DF1x										ɟ						

References

Peter Constable (2003) ‘Proposal to encode additional phonetic symbols in the UCS’, L2/03-190R.

John Laver (1994) *Principles of Phonetics*. Cambridge University Press.

Figure

This is the self-evident character for people who use <ʃ> and <ʒ> for Polish sz and ź (or for Russian ш and ж) along with the still-common IPA convention of representing affricates by ligatures, as seen for example in Figure 1. Figure 1 shows only the voiceless affricate /tʃ/, but Polish has a corresponding voiced affricate that needs to be transcribed the same way. Note that Polish contrasts affricates such as dź /ɟʒ/ with plosive-fricative sequences such as drz /dʒ/.

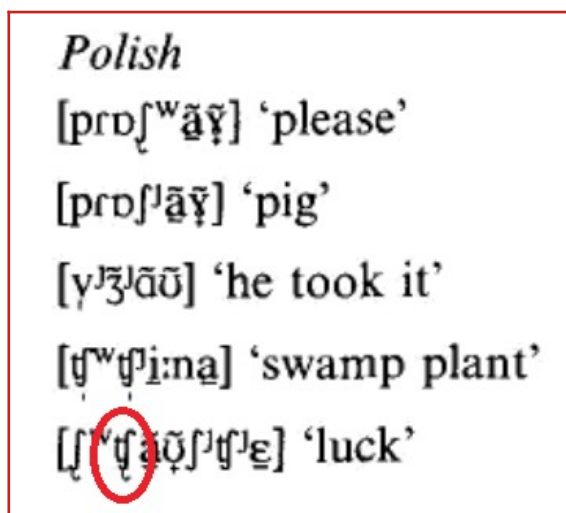


Figure 1. Laver (1994: 560) uses only one letter of the pair in this Polish example in his textbook on phonetics. Laver was unresponsive last year to queries on the need for both letters <ʃ ɟ> of the pair during the writing of the parent request to this proposal, and died in May 2020.

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 <http://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 <http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

A. Administrative

1. Title:	<input type="text"/>	<i>Additional phonetic click letters</i>
2. Requester's name:	<input type="text"/>	<i>Kirk Miller, Michael Everson</i>
3. Requester type (Member body/Liaison/Individual contribution):	<input type="text"/>	<i>individual</i>
4. Submission date:	<input type="text"/>	<i>2021 January 03</i>
5. Requester's reference (if applicable):	<input type="text"/>	
6. Choose one of the following:		
This is a complete proposal:	<input type="checkbox"/>	<i>yes</i>
(or) More information will be provided later:	<input type="checkbox"/>	

B. Technical - General

1. Choose one of the following:		
a. This proposal is for a new script (set of characters):	<input type="checkbox"/>	
Proposed name of script:	<input type="text"/>	
b. The proposal is for addition of character(s) to an existing block:	<input checked="" type="checkbox"/>	<i>yes</i>
Name of the existing block:	<input type="text"/>	<i>Latin Extended-G</i>
2. Number of characters in proposal:	<input type="text"/>	<i>1</i>
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"/>	B.2-Specialized (large collection) <input type="checkbox"/>
C-Major extinct <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?	<input checked="" type="checkbox"/>	<i>yes</i>
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	<input checked="" type="checkbox"/>	<i>yes</i>
b. Are the character shapes attached in a legible form suitable for review?	<input checked="" type="checkbox"/>	<i>yes</i>
5. Fonts related:		
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<input type="text"/>	
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<input type="text"/>	
6. References:		
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<input checked="" type="checkbox"/>	<i>yes</i>
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<input checked="" type="checkbox"/>	<i>yes</i>
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)?	<input checked="" type="checkbox"/>	<i>yes</i>

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 <http://www.unicode.org> for such information on other scripts. Also see Unicode Character Database (<http://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>yes</i> <i>John Esling, Michael Ashby</i> <i>(emails available on request, willing to write formal letters)</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>yes</i> <i>see illustrations</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>NA</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>yes</i> <i>(Unicode disprefers use of combining retroflex hook)</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: Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? If YES, reference:	<i>no</i> <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>