## **Revised Proposal to Encode Phonetic Symbols with Palatal Hook in the UCS**

*Date:* 2004-02-01

Author: Peter Constable, Microsoft

Address: One Microsoft Way

Redmond, WA 98052

**USA** 

*Tel:* +1 425 722 1867

Email: petercon@microsoft.com

| A.  | Administrative                   |  |
|-----|----------------------------------|--|
| 1.  | Title                            | Revised Proposal to Encode Phonetic Symbols with Palatal Hook in the UCS |
| 2.  | Requester's name                 | SIL International (contact: Jonathan Kew), Peter Constable               |
| 3.  | Requester type                   | Expert contribution  |
| 4.  | Submission date                  | 2004-02-01   |
| 5.  | Requester's reference            | L2/03-169  |
| 6a. | Completion                       | This is a complete proposal  |
| 6b. | More information to be provided? | Only as required for clarification.                                      |

| B.  | Technical—General                               |                           |
|-----|---|---------------------------|
| 1a. | New Script? Name?                               | No                        |
| 1b. | Addition of characters to existing block? Name? | Yes — Phonetic Extensions |
| 2.  | Number of characters in proposal                | 15                        |
| 3.  | Proposed category                               | A                         |

4. Proposed level of implementation 1 (no combining marks or jamo) and rationale 5a. Character names included in Yes proposal? Character names in accordance 5b. Yes with guidelines? Character shapes reviewable? Yes 5c. 6a. Who will provide computerized SIL International font? 6b. Font currently available? Yes Font format? TrueType 6c. 7a. Are references (to other character Yes sets, dictionaries, descriptive texts, etc.) provided? Are published examples (such as Yes samples from newspapers, magazines, or other sources) of use of proposed characters attached? 8. Does the proposal address other Yes, suggested character properties are included (see aspects of character data section E). processing?

### Technical—Justification

| <u>U.</u> | rechnical—Justification  |  |
|-----------|--|--|
| 1.        | Has this proposal for addition of character(s) been submitted before?      | An earlier proposal (L2/03-169) was reviewed by UTC (meeting #95), and concern was expressed at characters for which no evidence of attestation was provided. Such characters are documented here but have been removed from the inventory of characters being proposed. |
| 2a.       | Has contact been made to members of the user community?                    | No   |
| 2b.       | With whom?   | n/a  |
| 3.        | Information on the user community for the proposed characters is included? | Linguists specializing in Russian or Slavic languages.<br>Also used by some other linguists in relation to other languages.  |
| 4.        | The context of use for the proposed characters                             | Linguistic descriptions (books, journal publications, etc.); dictionaries.   |
| 5.        | Are the proposed characters in current use by the user community?          | Yes  |

Must the proposed characters be Preferably entirely in the BMP? 6b. Rationale? Living script / characters in current use 7. Should the proposed characters be Preferably kept together in a contiguous range? Can any of the proposed characters Possibly, but encoding of atomic characters is 8a. be considered a presentation form considered preferable (see discussion in section F of an existing character or character below) sequence? See discussion in section F below. 8b. Rationale for inclusion? 9a. Can any of the proposed characters No be considered to be similar (in appearance or function) to an existing character? 9b. Rationale for inclusion? n/a 10. Does the proposal include the use No. of combining characters and/or use of composite sequences? Does the proposal contain No. characters with any special properties?

#### D. SC2/WG2 Administrative

- Relevant SC2/WG2 document numbers
- Status (list of meeting number and corresponding action or disposition)
- Additional contact to user communities, liaison organizations, etc.
- 4. Assigned category and assigned priority/time frame

Other comments

#### E. Proposed Characters

A code chart and list of character names are shown on a new page.

#### **E.1** Code Chart

|   | хх0                 |
|---|---------------------|
| 0 | þ                   |
| 1 | d                   |
| 2 | f                   |
| 3 | g                   |
| 4 | ķ                   |
| 5 | 1                   |
| 6 | m                   |
| 7 | ŋ                   |
| 8 | p                   |
| 9 | ŗ                   |
| A | Ş                   |
| В | $\int_{\mathbb{F}}$ |
| С | Ų                   |
| D | X,                  |
| E | Z                   |
| F |                     |

#### **E.2** Character Names

| xx00 | LATIN SMALL LETTER B WITH PALATAL HOOK   |
|------|--|
| xx01 | LATIN SMALL LETTER D WITH PALATAL HOOK   |
| xx02 | LATIN SMALL LETTER F WITH PALATAL HOOK   |
| xx03 | LATIN SMALL LETTER G WITH PALATAL HOOK   |
| xx04 | LATIN SMALL LETTER K WITH PALATAL HOOK   |
| xx05 | LATIN SMALL LETTER L WITH PALATAL HOOK   |
| xx06 | LATIN SMALL LETTER M WITH PALATAL HOOK   |
| xx07 | LATIN SMALL LETTER N WITH PALATAL HOOK   |
| xx08 | LATIN SMALL LETTER P WITH PALATAL HOOK   |
| xx09 | LATIN SMALL LETTER R WITH PALATAL HOOK   |
| xx0A | LATIN SMALL LETTER S WITH PALATAL HOOK   |
| xx0B | LATIN SMALL LETTER ESH WITH PALATAL HOOK |
| xx0C | LATIN SMALL LETTER V WITH PALATAL HOOK   |
| xx0D | LATIN SMALL LETTER X WITH PALATAL HOOK   |
| xx0E | LATIN SMALL LETTER Z WITH PALATAL HOOK   |

#### **E.3 Unicode Character Properties**

All of these characters should have a general category of Ll; no case mapping for these characters is proposed. Other properties should match those of similar characters (e.g. U+01AB LATIN SMALL LETTER T WITH PALATAL HOOK).

#### F. Other Information

#### F.1 Background: transcription conventions for palatalization

In phonetic transcription, vowel symbols with palatal hook are generally used to represent consonants with palatalized articulation. Since 1989, the representation recommended by the International Phonetic Association has been to use superscript j; that is, the UCS character U+02B2 MODIFIER LETTER SMALL J.

Prior to 1989, however, IPA practice allowed for the use of palatal hook on consonant symbols. The older representation is still documented in the *IPA Handbook* (IPA 1999),<sup>1</sup> and is often referred to in general books on phonetics.

# OTHER LETTERS 28. Palatalised consonants, as in Russian: [t, d, n, etc. (see § 29 (i)). Special forms for palatalised f, 3 (for use when

Figure 1. From IPA (1949), p. 13.

**Secondary Articulations:** an opener articulation (usually of approximant type) superimposed on a simultaneous closer articulation.

Labialized. Simultaneous lip-rounding, e.g. [tw] [dw] [sw] [xw] etc.

Palatalized. Simultaneous raising of tongue dorsum towards the hard palate, e.g.  $[p^j]$   $[d^j]$   $[s^j]$  etc. also symbolized as [p] [d] [s] etc.

Figure 2. From Catford (1988), p. 222.

Within the linguistics tradition for study of the Russian language, use of characters with palatal hook has been common practice.

Revised Proposal to Encode Phonetic Symbols with Palatal Hook in the UCS Page 5 of 13
Peter G. Constable February 01, 2004 Rev: 23

<sup>&</sup>lt;sup>1</sup> Characters with palatal hook are not, in fact, used in that publication, however.

The palatalisation of consonants, one of the main phenomena in Russian speech habits, is shown in spelling in a very ingenious but indirect way, which is misleading and confusing for foreigners.

The palatalised consonants are not shown as palatalised but there are two sets of vowel letters.

I. ы а о у (i a o u) 2. и я е, ё ю (i ja jɛ, jo ju)

The letters of the first row are written after hard consonants and initially (except b)—

ды (di) да (da) до (do) ду (du)

The letters of the second row are written-

I. After soft consonants :

ди (di) дя (da) де, дё (de- or de-, do) дю (du) 2. After vowels and also initially:

моя́ ma'ja [my feminine], моё́ ma'jo [my neuter], мою́ ma'ju [my feminine accusative]; initially я ja [I], ел jel [I ate], е́ли 'jelı [we, you, they ate], е́лка 'jolkə [Christmas tree], юг juk [the South].

Figure 3. Consonants with palatal hook used for Russian (Boyanus and Jopson 1939, p. xxv).

specially designed letters. In this book we adopt largely the second method: palatalized consonants are indicated by a normal consonant letter with a small j attached to it at the lower right-hand corner. Thus, d represents 'palatalized d', represents 'palatalized s', g represents 'palatalized g' and so on.

Figure 4. Consonants with palatal hook used for Russian (Jones and Ward 1969, p. 82).

Characters with palatal hooks have been used in relation to other languages as well, however:

#### 744 Nick Evans

ed.ed.es "heavy rain" and alg.alg.al "straight as a ram-rod", "copy the second VC syllable", as in iy.alm.alm.ey "keeps playing", or "copy the third VC sylla-

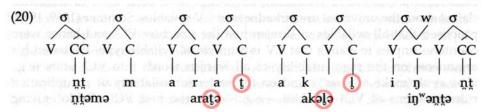


Figure 5. Examples of characters with palatal hook used in relation to Australian languages (Evans 1995, p. 744).

| Velar unvoiced    |    |      |   |          |             |
|-------------------|----|------|---|----------|-------------|
| fricative         | х  | x    | χ | χ        | x, x, g, kh |
| palatalized       | xy | (\$) |   | ź        |             |
| Palatal semivowel | у  | j    | у | y        | у           |
| Alveolar voiced   |    |      |   |          |             |
| fricative         | Z  | Z    | z | z        | Z           |
| [retroflex        |    | Z    | 7 | 7        | z ]         |
| palatalized       | zy | (3)  | ż | <b>ž</b> |             |
| labialized        | 7  |      |   | z        | zv, zw, ż   |

Figure 6. Examples of characters with palatal hook used in relation to African languages (Tucker 1971, p. 648).

It is in relation to Russian that the widest selection of symbols with palatal hook are used, however, and the inventory proposed here is based on the requirements for Russian. An inspection of a reasonably representative sampling of the linguistics literature suggests that this may be a complete inventory of required palatal-hook characters: apart from the characters proposed here and those already encoded in the UCS (e.g. U+01AB LATIN SMALL LETTER T WITH PALATAL HOOK), I have not found clear attestation of any other phonetic symbols using palatal hook. There is a pair of marginal cases, c and ezh, for which use of palatal-hook forms has not been clearly attested, but for which evidence indicates a need to encode palatal-hook forms may possibly arise in the future; these will be described below. Beyond these, however, I know of no additional candidates for palatal-hook forms.

Various authors have used typographic approximations for palatal hook when the selection of type available to them has not been extensive enough. This can be seen in Figure 6, above, in which a comma is used; others have used a cedilla:

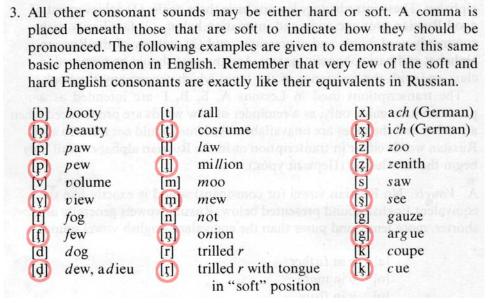


Figure 7. Cedilla as typographic approximation of palatal hook (Clark 1983, p. xx).

One other convention used by Slavicists is to indicate palatalization using a modifier letter apostrophe; e.g.,  $/t^2$ . A sample following this convention can be seen in Figure 12, below.

That approximations such as comma are used as a fallback when adequate type is not available can be seen in cases where conventions are mixed:

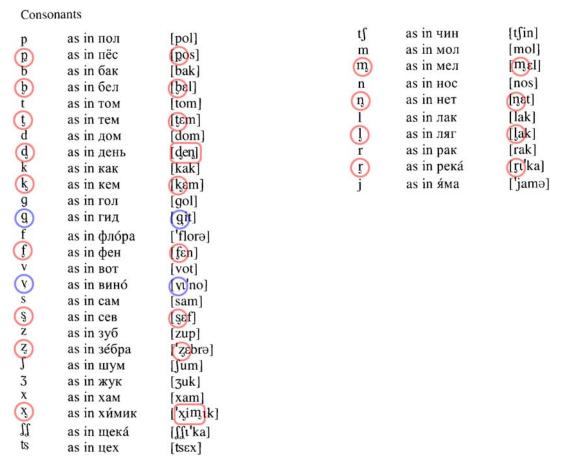


Figure 8. Comma as fallback approximation of palatal hook (Wade 2000, pp. 3-4).

In Figure 8, the use of true palatal-hook characters for the Russian palatalized consonants in all cases but g and v demonstrate clearly that this was the author's preferred practice for representing palatalization. And it is clear from other examples involving Russian (see, for instance Figure 7) that g-comma and v-comma are intended to represent palatalized consonants. We can only conclude that the author did not use palatal-hook typeforms in these two cases because they were not available to him.

#### F.2 Inventory of proposed palatal-hook characters

The inventory of proposed characters corresponds to palatalized consonant phones of Russian. The most commonly-encountered palatal-hook symbols can be seen in the sample from Jones and Ward (1969) shown in Figure 9:

#### Chart of Russian Consonant Phonemes Denti- Alveolar Post-Palatal Velar Bilabial Labio-Dental alveolar dental alveolar palatalized palatalized palatalized palatalized k Plosive t ţ P B Ь Ь d d g Fricative ſ X f f S ş Z 3 Y Z Affricate ts it[\* Nasal m m n ŋ 1 Lateral Rolled/Flapped r ŗ Semi-vowel j

Figure 9. Russian palatalized consonant phonemes (Jones and Ward 1969, p. 299).

This set of thirteen characters with palatal hook is consistently corroborated by several authors. (Note that one of these, t-palatal hook, is already encoded in the UCS. Hence, this accounts for twelve of the fifteen characters proposed.)

Other sources use additional characters with palatal hook in order to transcribe phonetic surface forms in Russian. Thus, the occurrence of palatal-hooked variants for g, esh, and x in Figure 11 below; the g-palatal hook can also be seen in Figure 4 above, and the x-palatal hook is also seen in Figure 10:

Advanced velar consonants, as in Russian: k, x. When, as in narrow transcription of Russian, separate letters are needed to denote ordinary g and advanced g, it is recommended that

Figure 10. Character x with palatal used for Russian (IPA 1949, p. 14).

| Greek  | Russian   | Italic  | Names of<br>Letters                      | Pronunciation  | Notes   |  |
|--|---|---|--|--|---|--|
| A<br>B   | Аа<br>Б6<br>Вв  | А а<br>Б б<br>В в   | a<br>be<br>ve                            | a, æ, ə, ı<br>b, b, p, p<br>v, y; f, f   | Modification of B cf. Sevastopol (for Sebastopol) Greek pronunciation         |  |
| Γ<br>Δ<br>E  | Гг<br>Дд<br>Ееё<br>Жж   | Г г<br>Д д<br>Е е ё<br>Ж ж  | ge<br>de<br>je<br>ze                     | g, g,; k<br>d, d; t, t<br>jɛ, je, ɛ, e; jo, o, ı<br>g; ∫   | Greek Greek  Modification of the corresponding "voiceless" Russian symbol III |  |
| Z<br>H   | Зз  | 3 3<br>И и  | zε<br>i                                  | z, z, s, s, i; I   | Modification of the Greek H  Other Names:  kpátkoe li kratkeje                |  |
|  | й   | ıй  | ij                                       | j  | [short i] li c краткой li s kratkej li with a sign of brevity]                |  |
| $K \wedge M$ $N \wedge O \cap P$ $X \wedge Y \wedge Y \wedge Y$ $X \wedge Y \wedge $ | К к<br>П л<br>М м<br>Н н<br>О о<br>П п<br>Р р<br>С с<br>Т У у<br>Ф Х х<br>Ц ц | К к<br>Л л<br>М м<br>Н н<br>О о<br>П п<br>Р р<br>С с<br>Т т<br>У у<br>Ф ф<br>Х х<br>Ц ц | ka el el en en o pe er es te u ef xa tse | k, k, l, l, m, m, m, m, m, n, n, o; a, ə P, P, p, r, r, s, s, t, t, t, u f, f, f, x, x, t, | Greek  (cf. π in Maths.)  Greek  Greek  |  |
|  | ШШ  | Ц и<br>Ш ш<br>Щ щ   | tfe<br>fa<br>f,tfa                       | î.<br>A  | Cf. Hebrew w  III+T with T under III, i.e. III  Look Other Names:             |  |

Figure 11. Palatal-hook characters used for Russian (Boyanus and Jopson 1939, p. xxiv).

The inventory from Boyanus and Jopson (1939) in Figure 11 with the exception of esh-palatal hook is corroborated by Ward (1966), by Clark (1983) (see Figure 7) and by Dawson et al (1964). This inventory is also corroborated by Wade (2000) (see Figure 8), though that author uses the IPA symbol esh-curl (U+0286) rather than esh-palatal hook.

#### F.3 Marginal cases: c, ezh

Some descriptions of Russian also make reference to palatalized post-alveolar voiced fricative and voiceless affricate, as shown in Figure 12:

Figure 12. Russian palatalized consonants, including post-alveolar fricatives and affricate (Halle 1994, p. 42).

Note that, in the chart in Figure 12, the author presents a complete inventory of palatalized consonants but is using the alternate convention of indicating palatalization by means of a modifier letter apostrophe, mentioned above. Also, this author is using the hacek diacritic to represent post-alveolar sounds:  $\S$ ,  $\S$  and  $\S$  rather than  $\S$ ,  $\S$  and  $\S$ . Thus, the palatalized post-alveolar voiced fricative and voiceless affricate are represented as  $\S$  and  $\S$  respectively. These phones are also attested using the comma representation described above:

```
Rule P 4 turns \[dožž+'a\] into \[dožž,+'a\]; \[n,'iščij\] into \[n,'iš,š,ij\]; and \[r,'ešči\] into \[r,'eš,š,i\].

Rule P 5b specifies \[&\] in \[s=&c'a*st,ju\] as sharped.

Rule P 6b turns \[s=&c,'a*st,ju\] into \[s,=&c,'a*st,ju\].

Rule P 7a turns \[s,=&c,'a*st,ju\] into \[s,=&c,'ast,ju\].

Rule P 8 turns \[dož,z,+'a\] into \[daž,z,+'a\].
```

Figure 13. Comma used as typographic approximation of palatal hook (Halle 1971, p. 52).

Again, it appears that, in such situations, the author has used a comma approximation of palatal-hook forms simply because adequate type that included characters with palatal hooks was not available.

The implication of this is that, were the type available, the author might have used c-palatal hook to represent the palatalized post-alveolar voiceless affricate. Also, with an author that used ezh rather than z-hacek for the voiced post-alveolar affricate, it seems possible that ezh-palatal hook might have been used to represent the palatalized variant of that sound.

Potential use of c-palatal hook is also suggested from the following sample from Africanist literature, in which c-comma is used for a palatalized consonant (Figure 6, repeated here for convenience as Figure 14):

| Velar unvoiced    |    |      |   |   |             |
|-------------------|----|------|---|---|-------------|
| fricative         | х  | x    | χ | X | x, x, g, kh |
| palatalized       | xy | (\$) |   | χ |             |
| Palatal semivowel | у  | j    | у | y | у           |
| Alveolar voiced   |    |      |   |   |             |
| fricative         | Z  | Z    | z | Z | Z           |
| [retroflex        |    | Z    | 7 | z | z ]         |
| palatalized       | zy | (3)  | ż | ž |             |
| labialized        | 7  |      |   | z | zv, zw, ż   |

Figure 14. C-comma used as typographic approximation for c-palatal hook (Tucker 1971, p. 648).

In this work, the author is presenting various representations for phones of sub-Saharan languages. The second column of his table is labeled "I.P.A.", and his practice in other cases of

palatalized consonants is to use a palatal-hook form, as seen here in the case of z-palatal hook. Thus, it appears that c-comma is being used due to a lack of type for c-palatal hook.

Therefore, in addition to those characters proposed here, there is evidence that suggests the possibility of eventually needing to represent c-palatal hook and z-palatal hook in the UCS. In the samples shown, however, various alternate representation conventions were used, and not the palatal-hook variants of c and z. In the absence of clear attestation for these characters, therefore, they are not included in this proposal. They are documented here, however, to show what the full extent of required palatal-hook characters might eventually be.

#### F.4 Representation as sequences with U+0321

Question 8a of section C above asks whether these characters can be considered presentation forms of existing characters or character sequences. They could possibly be viewed as sequences involving U+0321 COMBINING PALATALIZED HOOK BELOW. I suggest, however, that this would be inappropriate and is irrelevant. While combining marks in general are assumed to be applicable to arbitrary characters in a generative manner, allowing dynamic representation of text elements such as *Latin small a with bridge below*, there are certain combining marks for which this is not appropriate, one of these being U+0321 COMBINING PALATALIZED HOOK BELOW. This view has been expressed on the Unicore discussion list, and some of the rationale provided here has been expressed by others on that list.

There simply are only certain base characters that can sensibly be modified with a palatal hook, both in a linguistic sense as well as a typographic sense. For instance, it would be silly to encode a character sequence < U+01AB LATIN SMALL LETTER T WITH PALATAL HOOK, U+0321 COMBINING PALATALIZED HOOK BELOW >. In practice, there is a very limited inventory of characters that are used with palatal-hook modification.

Also, whereas it is feasible to create font/rendering implementations that can productively display sequences involving arbitrary base characters followed by a combining mark such as U+0300 COMBINING GRAVE ACCENT using mechanisms such as glyph attachment points, this is not feasible for U+0321 COMBINING PALATALIZED HOOK BELOW: the way in which a base character is modified using a palatal hook is dependent on the particular base character involved. Font implementations must assume a specific inventory of palatal-hook forms.

Thus, in terms of usage requirements and the realities of implementation, dynamic composition using U+0321 COMBINING PALATALIZED HOOK BELOW is not a good choice, and should be avoided.

Note that this view is corroborated by existing characters in Unicode itself in that characters such as U+01AB LATIN SMALL LETTER T WITH PALATAL HOOK do not have a decomposition. The combining mark U+0321 COMBINING PALATALIZED HOOK BELOW is not currently used in any decomposition, even though there are various potential candidates for such decompositions existing in the UCS.

Therefore, since there are good reasons why productive use of U+0321 COMBINING PALATALIZED HOOK BELOW is not recommended, and insofar as existing characters with palatal hook are not considered presentation forms of existing sequences, it is suggested that the characters proposed here are likewise not to be considered presentation forms of existing sequences.

#### G. References

- Boyanus, S.C., and N.B. Jopson. 1939. *Spoken Russian: A practical course.* London: Sidgewick and Sons.
- Catford, J.C. 1988. A practical introduction to phonetics. Oxford: Clarendon Press.
- Clark, Ben T. 1983. Russian. 3rd edn. New York: Harper & Row.
- Dawson, Clayton L.; Charles E. Bidwell; and Assya Humesky. 1964. *Modern Russian*. Washington, D.C.: Georgetown University Press.
- Evans, Nick. 1995. "Current issues in the phonology of Australian languages." *The handbook of phonological theory (Blackwell handbooks in linguistics,* 1), ed. by John A. Goldsmith, 723–61. Oxford: Blackwell Publishers.
- Halle, Morris. 1971. The sound pattern of Russian. The Hague: Mouton & Co.
- ——. 1994. "The Russian declension: an illustration of the theory of distributed morphology." *Perspectives in phonology (CSLI lecture notes,* 51), ed. by Jennifer Cole and Charles Kisseberth, 29–60. Stanford, CA: CSLI Publications.
- International Phonetic Association (L'association Phonétique Internationale). 1946. "parti administrati:v." *Le maître phonétique* 85:1, p. 16–7.
- ——. 1999. *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press.
- Jones, Daniel, and Dennis Ward. 1969. *The phonetics of Russian*. Cambridge: Cambridge University Press.
- Tucker, A.N. 1971. "Orthographic systems and conventions in Sub-Saharan Africa." *Current trends in linguistics, volume 7: Linguistics in Sub-Saharan Africa*, ed. by Thomas A. Sebeok, 618–53. The Hague: Mouton.
- Wade, Terence. 2000. A Comprehensive Russian Grammar. 2nd edn. Oxford: Blackwell.
- Ward, Dennis. 1966. Russian pronunciation illustrated. Cambridge: Cambridge University Press.