L2/03-169

Proposal to Encode Phonetic Symbols with Palatal Hook in the UCS

Date: 2003-5-30
Author: Peter Constable, SIL International
Address: 7500 W. Camp Wisdom Rd.
Dallas, TX 75236
USA
Tel: +1 972 708 7485
Email: peter_constable@sil.org

A. Administrative

1.	Title	Proposal to Encode Phonetic Symbols with Palatal Hook in the UCS
2.	Requester's name	SIL International (contact: Peter Constable)
3.	Requester type	Expert contribution
4.	Submission date	2003-05-30
5.	Requester's reference	
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	17
3.	Proposed category	A
4.	Proposed level of implementation and rationale	1 (no combining marks or jamo)
5a.	Character names included in proposal?	Yes
5b.	Character names in accordance with guidelines?	Yes
5c.	Character shapes reviewable?	Yes
6a.	Who will provide computerized font?	SIL International
6b.	Font currently available?	Yes
6c.	Font format?	TrueType

7a.	Are references (to other character sets, dictionaries, descriptive texts, etc.) provided?	Yes
7b.	Are published examples (such as samples from newspapers, magazines, or other sources) of use of proposed characters attached?	Yes
8.	Does the proposal address other aspects of character data processing?	Yes, suggested character properties are included (see section E).

C. Technical—Justification

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

SC2/WG2 Administrative

1. Relevant SC2/WG2 document numbers
2. Status (list of meeting number and corresponding action or disposition)
3. 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

	xx0	xx1
0	ƀ	ʒ̡
1	ɕ	
2	ɖ	
3	ɟ	
4	ɠ	
5	ƙ	
6	ɭ	
7	ɱ	
8	ɲ	
9	ƥ	
A	ɾ	
B	ʂ	
C	ʃ	
D	ʏ	
E	x	
F	ʐ	

E.2 Character Names

xx00 LATIN SMALL LETTER B WITH PALATAL HOOK
xx01 LATIN SMALL LETTER C WITH PALATAL HOOK
xx02 LATIN SMALL LETTER D WITH PALATAL HOOK
xx03 LATIN SMALL LETTER F WITH PALATAL HOOK
xx04 LATIN SMALL LETTER G WITH PALATAL HOOK
xx05 LATIN SMALL LETTER K WITH PALATAL HOOK
xx06 LATIN SMALL LETTER L WITH PALATAL HOOK
xx07 LATIN SMALL LETTER M WITH PALATAL HOOK
xx08 LATIN SMALL LETTER N WITH PALATAL HOOK
xx09 LATIN SMALL LETTER P WITH PALATAL HOOK
xx0A LATIN SMALL LETTER R WITH PALATAL HOOK
xx0B LATIN SMALL LETTER S WITH PALATAL HOOK
xx0C LATIN SMALL LETTER ESH WITH PALATAL HOOK
xx0D LATIN SMALL LETTER V WITH PALATAL HOOK
xx0E LATIN SMALL LETTER X WITH PALATAL HOOK
xx0F LATIN SMALL LETTER Z WITH PALATAL HOOK
xx10 LATIN SMALL LETTER EZH WITH PALATAL HOOK

E.3 Unicode Character Properties

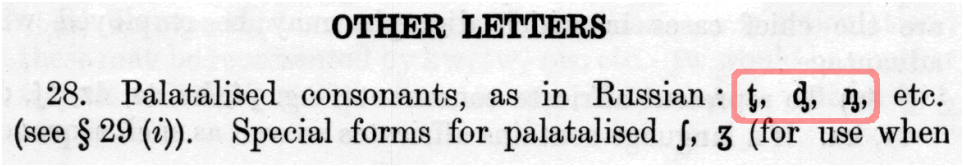
All of these characters should have a general category of Ll; no case mapping for these characters is proposed. Other properties should match those of similar characters (e.g. U+01AB LATIN SMALL LETTER T WITH PALATAL HOOK).

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.

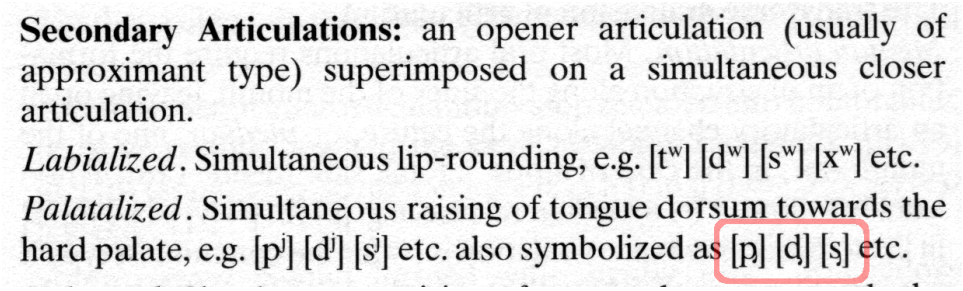
Prior to 1989, however, IPA practice allowed for the use of palatal hook on consonant symbols. The older representation is still documented in the *IPA Handbook* (IPA 1999),[1] and they are often referred to in general books on phonetics.



OTHER LETTERS

28. Palatalised consonants, as in Russian : ʈ, ɖ, ɳ, etc. (see § 29 (*i*)). Special forms for palatalised ʃ, ʒ (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.

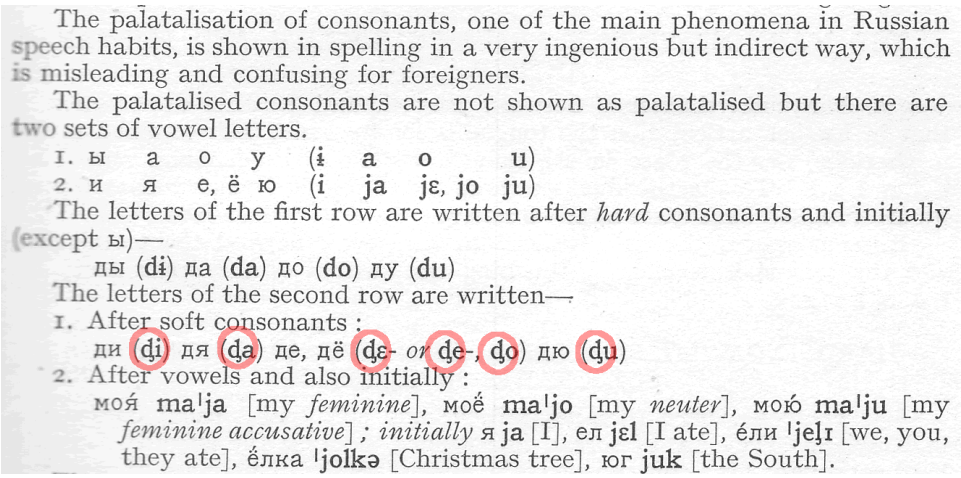
Labialized. Simultaneous lip-rounding, e.g. [tʷ] [dʷ] [sʷ] [xʷ] etc.

Palatalized. Simultaneous raising of tongue dorsum towards the hard palate, e.g. [pʲ] [dʲ] [sʲ] etc. also symbolized as [p̠] [d̠] [s̠] 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.

[1] 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.

1. ы а о у (ɨ 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 ы)—

ды (dɨ) да (da) до (do) ду (du)

The letters of the second row are written—

1. After soft consonants :

ди (dʲi) дя (dʲa) де, дё (dʲɛ- *or* dʲe-, dʲo) дю (dʲu)

2. After vowels and also initially :

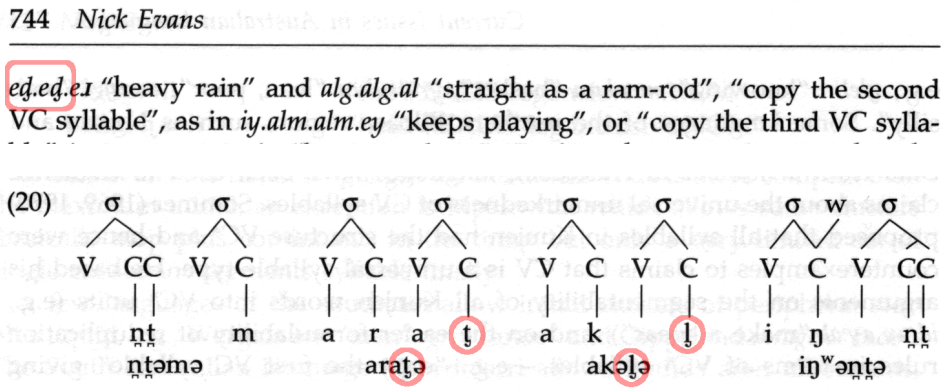
моя́ maˈja [my *feminine*], мое́ maˈjo [my *neuter*], мою́ maˈju [my *feminine accusative*] ; *initially* я ja [I], ел jɛl [I ate], е́ли ˈjeɭɪ [we, you, they ate], е́лка ˈjolkə [Christmas tree], юг juk [the South].

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, ɖ represents 'palatalized d', ʂ represents 'palatalized s', ɡ, 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:

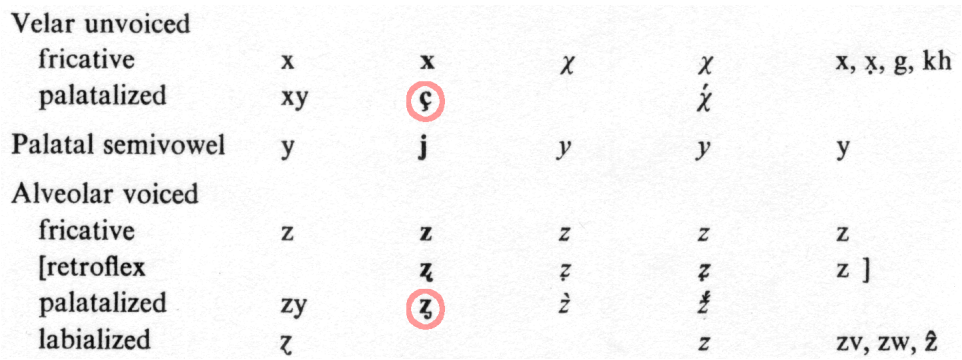


744 *Nick Evans*

eɖ.eɖ.eɪ "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-

(20)

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



Velar unvoiced					
fricative	x	x	χ	χ	x, x̣, g, kh
palatalized	xy	ɕ		χ́	
Palatal semivowel	y	j	*y*	*y*	y
Alveolar voiced					
fricative	z	z	*z*	*z*	z
[retroflex		ʐ	*ʐ*	*ʐ*	z]
palatalized	zy	ʒ	*ż*	*ʑ*	
labialized	ʐ			*z*	zv, zw, ẑ

Figure 6. Examples of characters with palatal hook used in relation to African languages (Tucker 1971, p. 648).[2]

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 is a complete inventory: 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 encountered any other phonetic symbols using palatal hook.

One other convention used by Slavicists is to indicate palatalization using a modifier letter apostrophe; e.g., /t’/. Also, some authors have typographically approximated characters with palatal hook using a comma; e.g., “t,”. Examples of each will be used in the discussion below.

F.2 Clearly-attested characters

Most of the inventory corresponds to palatalized consonant phonemes of Russian, as shown in the sample from Jones and Ward (1969) shown in Figure 7:

[2] Note that this author is using a comma for typographic approximation of c-palatal hook. See section F.3 for further discussion.

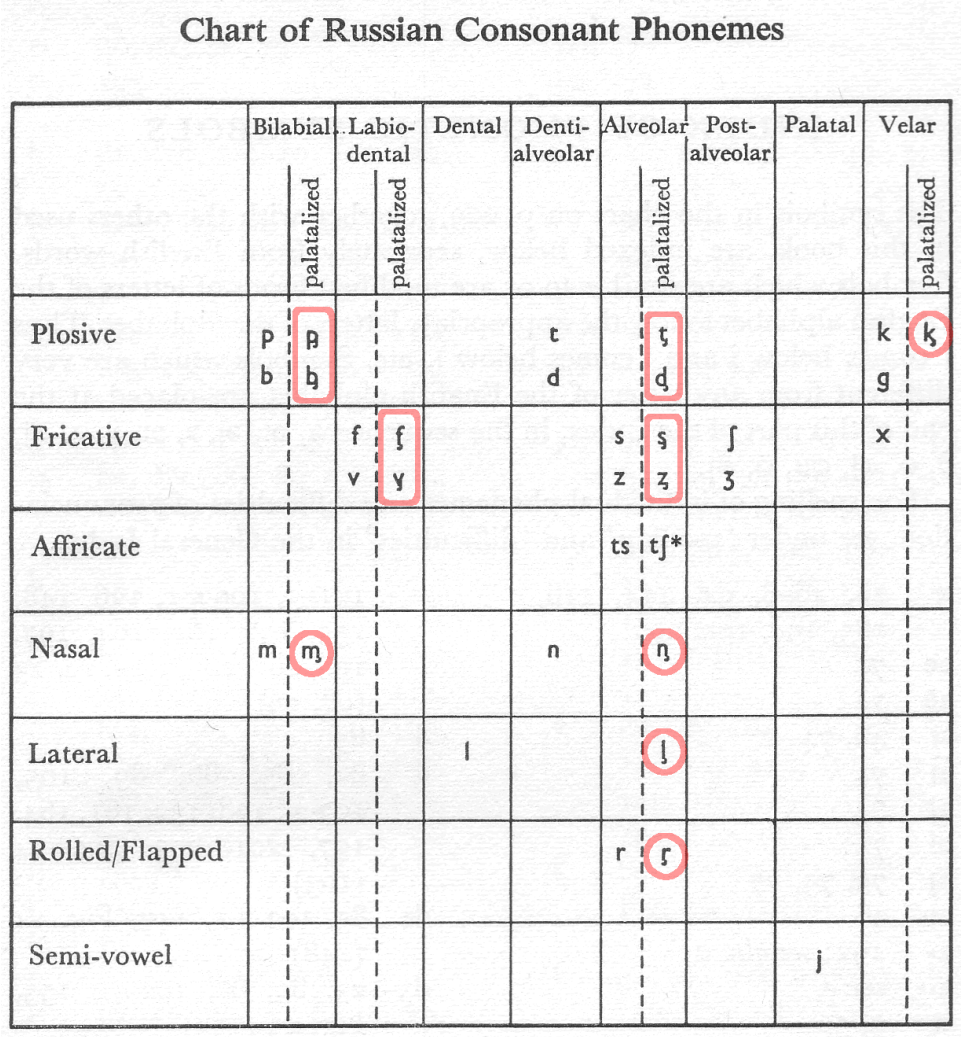
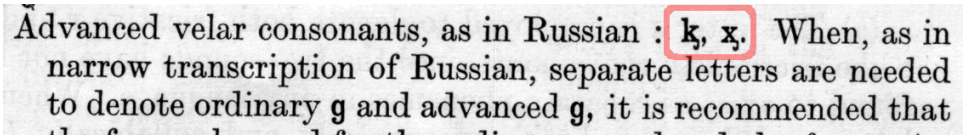


Chart of Russian Consonant Phonemes

	Bilabial		Labio-dental		Dental	Denti-alveolar	Alveolar		Post-alveolar	Palatal	Velar	
		palatalized		palatalized				palatalized				palatalized
Plosive	p b	ʙ ḇ				t d		ʈ ɖ			k g	ḳ
Fricative			f v	f̡ v̡			s z	ʂ ʐ	ʃ ʒ		x	
Affricate							ts	tʃ*				
Nasal	m	m̡				n		ɳ				
Lateral					l			ɭ				
Rolled/Flapped							r	ɾ				
Semi-vowel										j		

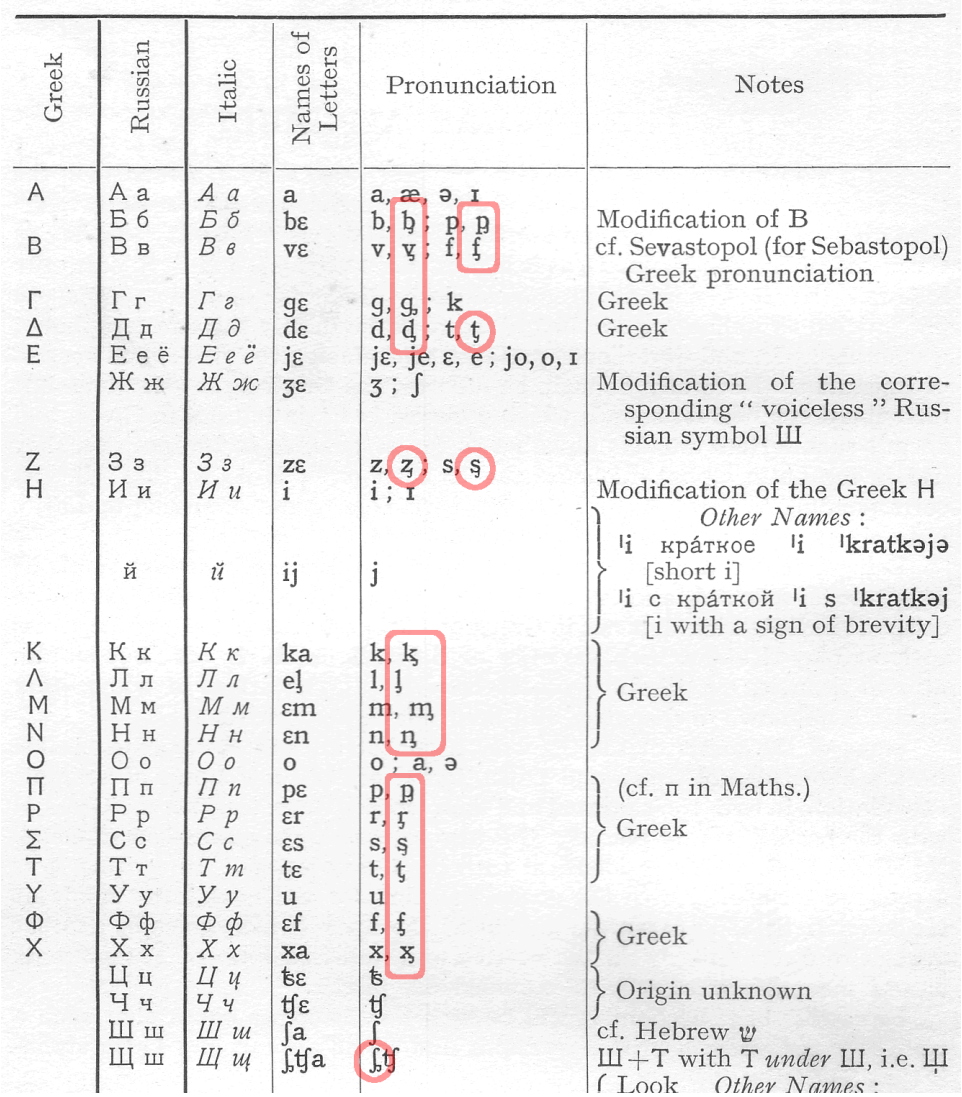
Figure 7. Russian palatalized consonant phonemes (Jones and Ward 1969, p. 299).

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 17 characters proposed.) Other sources use additional characters with palatal hook in order to transcribe phonetic surface forms. Thus, the occurrence of palatal-hooked variants for g, esh, and x in Figure 9 below; the g-palatal hook can also be seen in Figure 4 above, and the x-palatal hook is seen in Figure 8:



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

Figure 8. Character x with palatal used for Russian (IPA 1949, p. 14).



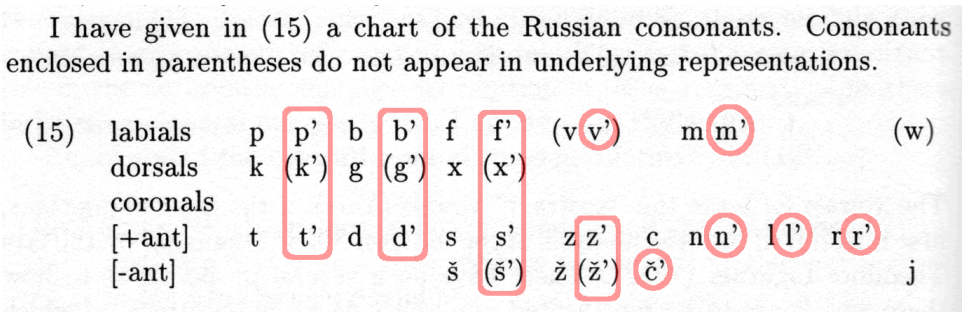
Greek	Russian	Italic	Names of Letters	Pronunciation	Notes
A	А а	*А а*	a	a, æ, ə, ɪ	
	Б б	*Б б*	bɛ	b, b̨ ; p, p̨	Modification of B
B	В в	*В в*	vɛ	v, v̨ ; f, f̨	cf. Sevastopol (for Sebastopol) Greek pronunciation
Γ	Г г	*Г г*	gɛ	g, g̨ ; k	Greek
Δ	Д д	*Д д*	dɛ	d, d̨ ; t, t̨	Greek
E	Е е ё	*Е е ё*	jɛ	jɛ, je, ɛ, e ; jo, o, ɪ	
	Ж ж	*Ж ж*	ʒɛ	ʒ ; ʃ	Modification of the corresponding " voiceless " Russian symbol Ш
Z	З з	*З з*	zɛ	z, z̨ ; s, s̨	
H	И и	*И и*	i	i ; ɪ	Modification of the Greek H
					Other Names : ˈi кра́ткое ˈi ˈkratkəjə [short i] ˈi с кра́ткой ˈi s ˈkratkəj [i with a sign of brevity]
	й	*й*	ij	j	
K	К к	*К к*	ka	k, k̨	
Λ	Л л	*Л л*	eļ	l, l̨	Greek
M	М м	*М м*	ɛm	m, m̨	
N	Н н	*Н н*	ɛn	n, n̨	
O	О о	*О о*	o	o ; a, ə	
Π	П п	*П п*	pɛ	p, p̨	(cf. π in Maths.)
P	Р р	*Р р*	ɛr	r, r̨	Greek
Σ	С с	*С с*	ɛs	s, s̨	
T	Т т	*Т т*	tɛ	t, t̨	
Y	У у	*У у*	u	u	
Φ	Ф ф	*Ф ф*	ɛf	f, f̨	Greek
X	Х х	*Х х*	xa	x, x̨	
	Ц ц	*Ц ц*	ʦɛ	ʦ	Origin unknown
	Ч ч	*Ч ч*	ʧɛ	ʧ	
	Ш ш	*Ш ш*	ʃa	ʃ	cf. Hebrew ש
	Щ ш	*Щ щ*	ʃ̡ʧa	ʃ̡ʧ	Ш+Т with Т *under* Ш, i.e. Щ
					(Look *Other Names* :

Figure 9. Palatal-hook characters used for Russian (Boyanus and Jopson 1939, p. xxiv).

The inventory from Boyanus and Jopson (1939) in Figure 9 with the exception of esh-palatal hook is corroborated by Ward (1966), by Clark (1983) and by Dawson et al (1964). Thus, there is clear attestation from the proposed characters other than c-palatal hook, esh-palatal hook and ezh-palatal hook. These last three will be discussed further.

F.3 Rationale for c-palatal hook, esh-palatal hook and ezh-palatal hook

The complete inventory of Russian palatalized phones includes palatalized voiced alveolar fricative and palatalized voiced alveolar affricate, as shown in Figure 10:

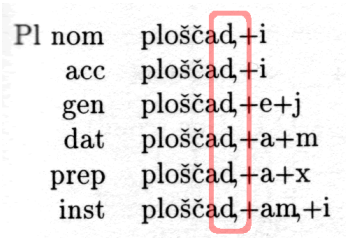


I have given in (15) a chart of the Russian consonants. Consonants enclosed in parentheses do not appear in underlying representations.

(15)	labials	p	p’	b	b’	f	f’	(v	v’)		m	m’							(w)
	dorsals	k	(k’)	g	(g’)	x	(x’)												
	coronals																		
	[+ant]	t	t’	d	d’	s	s’		z	z’		c		n	n’	l	l’	r	r’
	[-ant]					š	(š’)		ž	(ž’)		č’							j

Figure 10. Russian palatalized consonants, including alveolar fricatives and affricate (Halle 1994, p. 42).

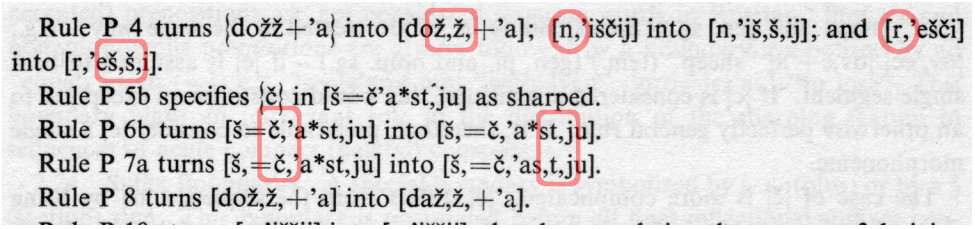
Note that, in the chart in Figure 10, the author is using a convention of indicating palatalization by means of a modifier letter apostrophe. Elsewhere in that article, he uses a comma as a typographic approximation for the palatal hook:



Pl nom	ploščad,+i
acc	ploščad,+i
gen	ploščad,+e+j
dat	ploščad,+a+m
prep	ploščad,+a+x
inst	ploščad,+am,+i

Figure 11. Comma used as typographic approximation of palatal hook (Halle 1994, p. 49).

This practice of using comma to approximate palatal hook is attested in other publications as well:



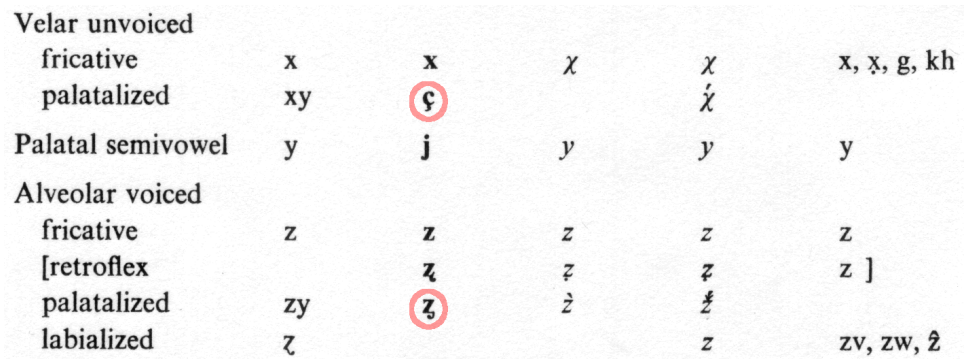
Rule P 4 turns {dožž+’a} into [dož,ž,+’a]; [n,’iščij] into [n,’iš,š,ij]; and [r,’ešči] into [r,’eš,š,j].
Rule P 5b specifies {č} in [š=č’a*st,ju] as sharped.
Rule P 6b turns [š=č,’a*st,ju] into [š,=č,’a*st,ju].
Rule P 7a turns [š,=č,’a*st,ju] into [š,=č,’as,t,ju].
Rule P 8 turns [dož,ž,+’a] into [daž,ž,+’a].

Figure 12. Comma used as typographic approximation of palatal hook (Halle 1971, p. 52).

It seems quite likely that the author has used a comma simply because adequate type that included characters with palatal hooks was not available.

Another point to be noted from Figure 10 and Figure 12 is the convention used by some Slavicists to transcribe alveopalatal fricatives using “š” and “ž” rather than esh and ezh, and to transcribe the voiceless alveopalatal affricate using “č” rather than “tʃ”. Note that Russian surface forms include *palatalized* voiceless alveolar affricate, and accordingly the symbols combined with comma in approximation of palatal hook in Figure 12 include “č”. It seems, then, that a c-palatal hook (in sequence with a combining caron) would be used by authors that follow the š/ž/č convention were that character available.

Other evidence for c-palatal hook has been encountered in Africanist literature, again using a typographic approximation:

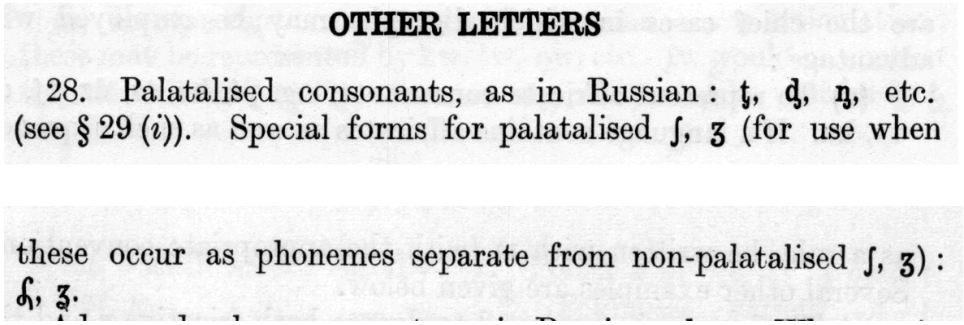


Velar unvoiced					
fricative	x	x	χ	χ	x, x̣, g, kh
palatalized	xy	ɕ		χ́	
Palatal semivowel	y	j	y	y	y
Alveolar voiced					
fricative	z	z	z	z	z
[retroflex		ʐ	ʐ	ʐ	z]
palatalized	zy	ʒ	ż	ź	
labialized	ẓ			z	zv, zw, ẑ

Figure 13. Comma used as typographic approximation for c-palatal hook (Tucker 1971, p. 648).

Note that this author does prefer to use true palatal-hooked characters, if available.

While the older IPA recommendation for palatalization involved several characters with the palatal hook, distinct symbols without palatal hook are recommended in IPA 1949 for the alveolar fricatives rather than esh-palatal hook and ezh-palatal hook:



OTHER LETTERS

28. Palatalised consonants, as in Russian: ʈ, ɖ, ɳ, etc. (see § 29 (*i*)). Special forms for palatalised ʃ, ʒ (for use when these occur as phonemes separate from non-palatalised ʃ, ʒ): ʃ̠, ʒ̠.

Figure 14. IPA recommendation: distinct symbols for palatal esh, ezh (IPA 1949, pp. 13–14).

Even so, it is known that at least esh-palatal hook has been used by some authors, as shown in Figure 9 above. It seems likely that, if such authors were to transcribe the voiced counterpart, they would want to use an ezh-palatal hook. Thus, even though other conventions for transcribing palatalized alveolar fricatives may be more common, the potential demand for esh-palatal hook and ezh-palatal hook is real.

Therefore, in spite of limited attestation, it seems likely that the characters c-palatal hook, esh palatal hook and ezh-palatal hook would be used by authors if available, in preference to typographic approximations seen in existing literature. By including these three characters along with the others proposed here, a complete inventory of characters with palatal hook that are ever likely to be required for phonetic transcription of Russian will have been provided in the UCS.

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 character or character sequences. They could possibly be viewed this way: 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 than 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.

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, though there are various potential candidates for such decompositions.

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.

Ward, Dennis. 1966. *Russian pronunciation illustrated.* Cambridge: Cambridge University Press.