Proposal to Encode Phonetic Symbols with Retroflex 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 Retroflex 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	9
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.
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?	These were more often used several decades ago, though some are attested in recent publications.
6a.	Must the proposed characters be entirely in the BMP?	Preferably
6b.	Rationale?	If possible, should be kept with other phonetic symbols in the BMP.
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.

D. 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.

Code Chart

	xx0
0	a̢
1	ɑ̢
2	e̢
3	ɛ̢
4	ɝ̢
5	ə̢
6	i̢
7	ɔ̢
8	u̢
9	
A	
B	
C	
D	
E	
F	

Character Names

xx00   LATIN SMALL LETTER A WITH RETROFLEX HOOK
xx01   LATIN SMALL LETTER ALPHA WITH RETROFLEX HOOK
xx02   LATIN SMALL LETTER E WITH RETROFLEX HOOK
xx03   LATIN SMALL LETTER OPEN E WITH RETROFLEX HOOK
xx04   LATIN SMALL LETTER REVERSED OPEN E WITH RETROFLEX HOOK
xx05   LATIN SMALL LETTER SCHWA WITH RETROFLEX HOOK
xx06   LATIN SMALL LETTER I WITH RETROFLEX HOOK
xx07   LATIN SMALL LETTER OPEN O WITH RETROFLEX HOOK
xx08   LATIN SMALL LETTER U WITH RETROFLEX HOOK

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+0273 LATIN SMALL LETTER N WITH RETROFLEX HOOK).

F.   Other Information

Rationale

In phonetic transcription, vowel symbols with retroflex hook are generally used to represent vowel phones with rhoticity ("r-colouring"). Since 1989, the representation recommended by the International Phonetic Association has been to use the rhotic hook; that is, the UCS characters U+025A LATIN SMALL LETTER SCHWA WITH HOOK and U+025D LATIN SMALL LETTER REVERSED OPEN E WITH HOOK, and otherwise a character sequence of a vowel sign followed by U+02DE MODIFIER LETTER RHOTIC HOOK.

Prior to 1989, however, IPA practice was to use a retroflex hook on vowel symbols. The older representation is still cited in the IPA Handbook (IPA 1999):

Figure 1. Samples of symbols with retroflex hook: IPA (1999), p. 173.

Vowel symbols with retroflex hook are still occasionally used by linguists in current publications, as seen in Figure 2:
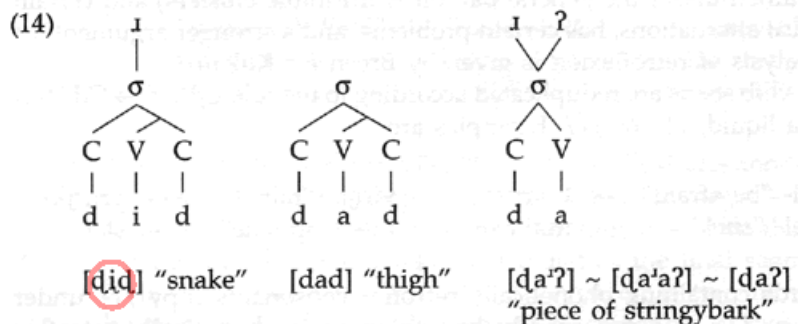

Figure 2. Latin small i with retroflex hook: Evans (1995), p. 740.

Current publications may also use these characters for purposes of citing historic practice, as illustrated in Figure 1.

Insofar as the current IPA recommendation is to use rhotic hook, it is suggested that the `NamesList.txt` file in the Unicode Character Database include an annotation to that effect.

The inventory of characters proposed is that which were approved by the International Phonetic Association in 1946, as shown in the following figures:

Figure 3. IPA vowel symbols with retroflex hook: IPA (1946), p. 16.

Figure 4. IPA vowel symbols with retroflex hook: IPA (1946), p. 16.

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 already encoded in the UCS (e.g. U+0290 LATIN SMALL LETTER Z WITH RETROFLEX HOOK), I have not encountered any other phonetic symbols using retroflex hook, except for the lone instance of inverted small-capital r with retroflex hook shown in Figure 1, which I take to be anomalous.

Representation as sequences with U+0322

Question 8a of section C above asks whether these characters can be considered a presentation form of an existing character or character sequence. They could possibly be viewed this way, but I suggest 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+0322 COMBINING RETROFLEX 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 retroflex hook, both in a linguistic sense as well as a typographic sense. For instance, it would be silly for both linguistic and typographic reasons to encode a character sequence < U+0290 LATIN SMALL LETTER Z WITH RETROFLEX HOOK, U+0322 COMBINING RETROFLEX HOOK BELOW >. In practice, there is a very limited inventory of characters that are used with retroflex-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+0322 COMBINING RETROFLEX HOOK BELOW: the way in which a base character is modified using a retroflex hook is dependent on the particular base character involved.

Thus, in terms of usage requirements and the realities of implementation, dynamic composition using U+0322 COMBINING RETROFLEX 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+0290 LATIN SMALL LETTER Z WITH RETROFLEX HOOK do not have a decomposition. The combining mark U+0322 COMBINING RETROFLEX HOOK BELOW is not currently used in any decomposition, though there are a number of potential candidates for such decompositions.

Therefore, since there are good reasons why productive use of U+0322 COMBINING RETROFLEX HOOK BELOW is not recommended, and insofar as existing characters with retroflex 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

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.

International Phonetic Association (L'association Phonétique Internationale). 1946. "parti administratiːv." *Le maître phonétique* 85:1, p. 16–7.

——. 1999. *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press.