

To: UTC

L2/09-063

From: Deborah Anderson, SEI

RE: Dialect-Specific Characters in Batak (Batak proposal: L2/08-011 WG2 N3320R)

Date: 29 January 2009

At the November 2008 Unicode Technical Committee, a question was raised regarding the apparent encoding of dialect-specific letters in the Batak script, and whether they might be better covered by glyphs in fonts. I followed up on this, and posed a number of questions for Uli Kozok, the Batak expert, contained in the email message below.

According to Prof. Kozok, one font will be created containing the entire Unicode Batak repertoire. He advises against handling dialect-specific letters as glyph variants within fonts, as there is some dialect overlap and ambiguity in manuscripts and encoding the dialect-specific letters as characters would assist philologists. Encoding the dialect letters separately would make it easier to implement Batak.

-----Original Message-----

From: kozokuni@gmail.com On Behalf Of Uli Kozok

Sent: Tuesday, October 07, 2008 12:40 PM

To: Deborah W. Anderson; Michael Everson

Subject: Re: Question on Batak characters

Hi Deborah, my apologies for the delay.

On Tue, Sep 30, 2008 at 3:26 PM, Deborah W. Anderson <dwanders@sonic.net> wrote:

> Dear Uli,

>

> 1a. I have a question about the encoding of certain dialect-specific letters in the proposal.

>

> In the latest version of the proposal,

> <http://std.dkuug.dk/jtc1/sc2/wg2/docs/n3320.pdf>, there is BATAK LETTER

> A as well as a BATAK LETTER SIMALUNGUN A and other dialect-specific letters.

>

> I recall you had explained that there is a strong feeling of ethnic

> identity among the various dialect groups, and I assume this is one

> reason for encoding the SIMALUNGEN (/MANDAILING/KAO/PAKPAK) letters separately.

>

> If one were to write a BATAK YA in a Simalungun-dialect word, would a

> Simalungun user object and claim it is incorrect, and say only the

> BATAK LETTER SIMALUNGUN YA is right? Do users see the dialect

> variations in HA, A, PA, NA, WA, GA, RA, MA, TA, SA, YA and LA as

> being on the spectrum of acceptable glyph variants of the letters?

The Simalungun letter YA is less of a variant in a visual sense, but it is distinctive and part of the Simulungun "orthographic package. When it comes to HA, PA, WA GA, LA, and especially MA and RA then users would without doubt object when a different form of the letter would be used. Besides, Simalungun also has a very distinctive SA which is much more frequently used than the other Batak variant.

- > I'm wondering if some of these differences in the letters, which
- > appear to in several cases rather minor (such as for YA), aren't
- > actually just glyph differences that could be covered in a font?

See above. I would advise against it.

- > I know you had mentioned some
- > minor font differences that could be covered by a font mechanisms,
- > such as OpenType features. Did you mean that OpenType features would
- > be acceptable for reflecting YA, A, PA, etc. in the various dialects,
- > or were you referring to other differences in the letters?

I was actually referring to yet other variants that are included in a separate font set in which I included genuine font/glyph variants.

- > Will there be a KARO font with the preferred shapes, as well as a
- > PAKPAK font, a SIMALUNGUN font, etc.?

That's what we currently have in our 8-bit fonts, but the Unicode character set we are proposing is a superset and only one font will be needed. This is better for the philologist as in the Batak manuscript tradition the borders between the 5 different fonts are not always clear. We have, for instance, manuscripts written by Karos who studied in Toba.

- > The advantage to covering the minor variations in the font is that
- > users could search for a particular Batak word with BATAK LETTER A and
- > it would be found if it contained a BATAK LETTER A, for example.
- >
- > In the current proposal, if a word in a document was spelled with
- > BATAK LETTER SIMALUNGUN A, and someone had typed in BATAK LETTER A in
- > the search box instead, it would not be found.
- >
- > Of course, using BATAK LETTER A would not guarantee that the user
- > would get the preferred letter shape (unless he/she is using a font
- > with the shapes or OpenType features to show the alternate glyphs).
- >
- > I need your input on this, as it affects the proposal and collation
- > for Batak. As it stands (at least for collation), Karo, Pakpak,
- > Simalungun, Toba, and Mandailing alphabets are very close to being 5
- > different scripts, rather than one Batak script.

Yes, the user clearly perceives them as different script although when we look at the scripts more closely, and especially at the local variants as found in traditional manuscripts then it can occasionally be quite difficult to tell whether that particular script is Toba, Mandailing, or Pakpak, for instance as they share a lot of common characters and as there is considerable overlap - e.g. although there is a Simalungun sa, they occasionally also use the "standard" Batak sa.

>

> 1b: The proposal has two forms of TOBA TA in the Toba alphabet. Why  
> are there two forms? (These correspond to BATAK LETTER SOUTHERN TA and  
> BATAK LETTER NORTHERN TA.)

In manuscripts originating from Toba we find both forms in about equal distribution. I named the one the southern TA as it is the only form of TA found south of Toba (Mandailing) whereas north of Toba only the northern form is used. Historically it seems that the southern TA was predominant in eastern Toba whereas the northern TA prevailed in western Toba.

Hope this helps,  
best regards,

Uli

--

Dr. Uli Kozok  
Associate Professor  
Indonesian-Malay Language Program  
2540 Maile Way, Spalding 255  
Honolulu, HI 96822, USA  
Tel: +1.808. 956 7574  
Fax: +1.808. 956 5978  
<http://www.hawaii.edu/indolang>  
<http://bahasa.net/online>