To: Script Ad Hoc
From: Jim Brase
Date: November 3 2021
Subject: Comments on "Updated proposal to encode the Tai Don script in UCS"

The following are comments sent in by Jim Brase, author of the Tai Viet proposal, on the "Updated proposal to encode the Tai Don script in UCS" by Kushim Jiang.

I Introduction

People using Tai Don script are mainly distributed in Jinping County, so Chinese literature usually refers to it as Jinping Tai script

[Brase] Do we have any data to support the claim that the majority of the script users are in Jinping Count (China)? With respect to Tai Don speakers, the Ethnologue lists 280,000 in Vietnam, 200,000 in Laos, and 15,000 in China. But I have no information on literacy rates in the Tai script for any of these groups

The Tai Viet script inherited many forms from the Tai Don script in its construction process, but the incompatible differences between the two make the Tai Don script need to be encoded separately. A detailed description can be found in [Brase, 2020] and in general:

There are phonological differences (more consonants and less vowels in Tai Don script);

There are irreconcilable phoneme-grapheme correspondences (Viet 𖹃 /m$_4$/ ~ Don 𖹃 /m$_2$/ ~ Viet 𖹃 /p$_4$/ ~ Don 𖹃 /p$_2$/ ~ Viet 𖹃 /f$_4$/ ~ Don 𖹃 /f$_2$/).

[Brase] If the scripts are unified, a Tai Don speaker would use:
- TAI VIET LETTER HIGH PO for Tai Don /m2/;
- TAI VIET LETTER HIGH FO for Tai Don /p2/;
and would require a new character to be encoded for Tai Don /f2/.
I believe there are one or two other sets of characters in which the phonemic values are similarly skewed.

I believe that such a solution would be workable, but it is complicated by the fact that there are more variations in the way Tai Don is written than there are in Tai Dam. I've lacked resources to determine which dialect of Tai Don is most central, and how many variations should be included.

Although I have always expected that the Tai Dam and Tai Don (and maybe even the Tai Daeng) would be unified, I can see two advantages in the current proposal:
1) It selects a default version of the Tai Don script--the Jinping version due to the fact that that is what the proposers have been working with. I'm not sure that that is the most central "script

dialect", but it does seem to be the one in which the most active development work is taking place.

2) In the event that additional characters are needed in the future for one of the script dialects in NW Vietnam or NE Laos, the Jinping dialect will be a closer starting point to work with than the Tai Dam dialect of Son La province. And the final product will be less complex than a single block that includes Tai Dam, Jinping Dai, and the variations of Lai Chau province.

The function of vowel-like characters is uncertain for the time being; if a consonant-based perspective is taken, then vowel-like characters should be treated as marks, which may be inconsistent with xx02 qo. Any discussion of this would be welcome.

[Brase] I would think that if you want all vowels to be treated as combining marks, then you would need another character that was similar to xx02 in appearance, but had the properties of a combining mark.

In Tai Viet, we chose to use visual order because everyone we talked to insisted that they typed text in visual order. Using visual order is also consistent with Lao.
If the users prefer to enter text in logical order, then you would use logical order.

2. Script Details

There are also two rules in Tai Don scripts: ☐ When xx31 i, xx32 ee, xx35 u, xx36 e, xx37 ue, or xx3d u appears in C[P]VT as V, the V above C will be transferred above T;

$$<n\ (k),\ \overset{\prime}{\circ}\ (ue),\ \overset{}{w}\ (m)> = n\overset{}{\overset{\prime}{w}}\ (kuem).$$

[Brase] This is an interesting rule. We did not include any rule of this nature in Tai Viet, but we could have. In hand written text, which served as a model for SIL's Tai Heritage font, there is great variation in where combining vowels are placed. When a syllable contained both and initial and final consonant, the vowel could be placed over either consonant, but my impression was that half way between the two was preferred--providing, of course, that there was room there.

As the function of the vowel-like characters and the encoding model are to be determined, the rules are also to be determined. See Chart 3 for decision making.

[Brase] I think there is one other feature of the language that comes to my mind that I don't see covered here. In Tai Dam, velar consonants can be labialized. Thus: /ka:j^4/ 'to scratch (an itch)'; /k^{w}a:j^4/ 'water buffalo'.

In the Tai Viet script, the labialization is written with a HIGH VO after the consonant. The HIGH VO becomes the base character for any vowel or tone marks.

I am traveling, and do not have any Tai Don references at hand as I write this. But if my memory is correct, Tai Don has this same feature. If so, make sure the encoding will handle it, and especially how it will behave with respect to combining marks and left side vowels if logical order is used.

3. Comparison to Tai Viet Script

Code Point	Tai Viet Glyph	Tai Viet Name	Value	Code Point	Tai Dón Glyph	Tai Dón Name	Value
		Chart 7 Comparison between Tai Viet script and Tai Don script					
AA80	ɳ	LOW KO	/k$_1$/	XX08	n	HIGH KO	/k$_1$/

[Brase] I note that the names for the high series and low series consonants are consistent with the Lao, Thai, and general usage in the linguistic community.

We deliberately reversed these in TAI VIET, at the request of the Tai people on the committee, because of their perception of which tones class was high and low.
Just pointing this out to ensure that it is checked and this is what is wanted. I always found that this was a detail that was easy to overlook.