L2/09-283

Doc Type: Partially documented proposal Title: Proposal to include Chinook Pipa script in UCS Source: Van Anderson Status: For public review. Replaces: <u>Proposal v. 4.1</u> Action: For review by community. Comments to <u>Van Anderson</u> Date: 2009-07-24 Discussion list: <u>Chinook in the UCS</u>

Overview

The Chinook Pipa script, an adaptation by father Jean Marie Raphael LeJeune of the Duployéan shorthand, is an historic script used for writing the Chinook Jargon and other languages of interior British Columbia. Its original use and greatest surviving attestation is from the run of the *Kamloops Wawa*, a (mostly) Chinook Jargon newsletter of the Catholic diocese of Kamloops, British Columbia, published 1891-1923. At the time, the Chinook Jargon trade language was spoken in an area encompassing SE Alaska, most of British Columbia, Washington State, western Montana, Oregon, Idaho, far northern California, and communities further afield. Although the Chinook Jargon was the lingua franca in many of these communities, it was generally a spoken, rather than written language. Most attempts at documentation used the Latin script to approximate Jargon phonology, and indeed, dictionaries of the Chinook Jargon usage, during the height of its spread and utility. There currently exists no formal encoding, in any context, for the representation of the Chinook Jargon. Indeed, the submission of the Chinook Pipa script to UCS has necessitated the creation, from scratch, of the first Chinook typeface, such an effort currently under way with glyph images available for review.

Structure

Chinook Pipa contains several classes of letters, differentiated by visual form - hence script function - and phonetic value. Letter classes include the line and arc consonants, circle vowels (A and the O/W vowels), nasal vowels, arc vowels, and non-conjoiners. Since Chinook Pipa is an adaptation of a shorthand system, strings of letters are intended, generally, to join together cursively to form nominally syllabic units. Many Pipa letters have variant forms, including compound letters and overlapping concatenated behaviors for initialisms and abbreviations. Excepting for reverse stroke direction of a few letters, Chinook Pipa is written LTR, quasi-syllable by syllable, in horizontal lines proceeding down the page, as with most European scripts.

Ordering

Ordering of the characters in the Chinook Pipa is undefined - the only lexicon using the script cites nominally in Latin alphabetical order - so allocation order in the Chinook Pipa Character Block is revisable up to inclusion in the standard. Essentially, a Unicode Standard that includes a Chinook Pipa Character Block will be the only official ordering of the script. The currently proposed allocation ordering and its basis is as follows: According to Father LeJeune's <u>Chinook</u> <u>Rudiments</u>, the characters encoded x00-x09 (P,T,F,K,L,M,N,Sh,S,O) double as the numbers 1-9&0. x09 - x0C (O,A,Oo,Ow) constitute the next basic vowels given in his introduction, with W+ forms 16 code points later. x0D (I) is another simple vowel with a W+ form 16 code points later and a variant form 32 code points later. x0E rounds out the basic vowels given in LeJeune's repertoire, while x0F (H) is a non-conjoining simple consonant with a Salish variant 16 code points later. The second column of the allocation begins (x10-x14; B,D,V,G,R) with the voiced counterparts (elongated form) of the first five consonants (x00-x04). x15 is reserved for a form of x05, and x16-x18 (Ng,Ch,Ts) are the dot modified forms of x06-x08. x19-x1D (Wo,Wa,Woo,Wow,We) are the W+ forms of the vowel sequence in column 1: x09-x0D. x1E is reserved, ostensibly, for another W+ vowel variant. Rounding out the second column is the Salish complement to 'H', ie x1F (X). The third column is generally given to compound consonants and vowel variants,

with two code points (x20&x21) reserved for theroetical 'P' and 'B' compounds. x22-x24 (Th,Dh,Kh) are 'T','D',&'K' variants. x25 is for a putative 'G' compound, while x26 & x27 (Lh,Rhh) are 'L' and 'R' compounds. x28-x2A are reserved for consonant compounds, and x2B (K') and x2C (hL) are the K/G and L/R compounds. x2D (E) is a variant of x0D sharing a shape and properties, but having a different orientation in many conjoining environments. x2E (Uh) is the dot modified variant of x0E and x0F is a reserved codepoint. In the last column come the Nasal Vowels, x30-x33 (An,In,On,Un) that only intermittently appear in the Wawa texts, but are neither composed characters nor variants. Next is the logograph /likalisti/ meaning *eucharist*

at x24 with the code points x25-x2D again reserved, this time for any other logographs encountered in Pipa texts and any compound vowels or vowel variants unearthed through scholarship into private Pipa texts. Ending the allocation are the Chinook Pipa Full Stop (x3E) and the Virama-like Chinook Pipa Concatenator (x3F) which encodes for abbreviations and similar constructions in the script.

Naming

The Chinook Pipa script has gone by many names. The most common of these are "Chinook Pipa" and "Wawa Writing". The former, meaning *Chinook writing*

in Chinook Jargon was used both as a Chinook Jargon and English name for the script. The latter was generally used only in an English context. It has been suggested that the name of the script could simply be "Chinook", and there is some merit, considering the redundancy of the phrase "Chinook Pipa script".

Some discussion has also centered around the glyph names having "old" phonetic values. Specifically, the phonemes represented by the characters "Oo" (x1A) and "U" (x1E) are found in modern orthography as "U" and "Yu" respectively. Given that the transliterations found in the source materials for the script, as well as in most historic lexica, and even used in modern place names, all converge on the historic (Oo/U) transliteration, it is felt that the given character names are the most transparent.

The character CHINOOK PIPA LETTER KK is used to represent the "crossed K", thought to indicate a glotallized K' sound. Likewise CHINOOK PIPA LETTER DH indicates the voiced *Th* sound.

Alphabetization

No information is available on alphabetization, as the dictionary portions of the <u>Chinook Rudiments text</u> are given in roughly Latin alphabetical order. <u>Other sources</u>

group words by novel alphabetization, no more or less canonical than any other. The most logical ordering, given the structure of the script, would be along the lines of P, B, T (Th), D (Dh), F, V, K (Kh, K'), G, L (Lh, hL), R (Rhh), M, N (Ng), Sh (Ch), S (Ts), O (Wo), A (Wa), I, E (We), Oo (Woo), Ow (Wow), U (Uh), H (X) then An, In, On, and Un. Given that alphabetization is not a defined property of the Unicode Standard, it would seem that the above or simple binary order would more than suffice for any implementation needing an order of alphabetization.

Chinook Pipa Codechart v.4.2									
Previous Version									
Glyphs					Character List				
	U+x00	U+x10	U+x20	U+x30	Arc Consonants	Non-conjoining Consonant			
0				1	05 (CHINOOK PIPA LETTER M • number 6	1F f CHINOOK PIPA LETTER X · Voiceless velar/uvular fricative			
						Compound Consonants			

$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$			ヽ ノ 、 ⊕	06) CHINOOK PIPA LETTER N • number 720 <reserved>07 → CHINOOK PIPA LETTER SH • number 821 <reserved>08 → CHINOOK PIPA LETTER S • number 923 → CHINOOK PIPA LETTER DH 23 → CHINOOK PIPA LETTER KH08 → CHINOOK PIPA LETTER S • number 925 <reserved>26 ★ CHINOOK PIPA LETTER LH 27 ★ CHINOOK PIPA LETTER RHH 28 <reserved>Simple Vowels 09 ○ CHINOOK PIPA LETTER O20 09 ○ CHINOOK PIPA LETTER O20</br></reserved></br></br></reserved></reserved></reserved>
$\begin{pmatrix} 6 \end{pmatrix}$) ()	•		 number 0 Circle vowel OA O CHINOOK PIPA CHINOOK PIPA
	$\dot{\cdot}$	<i>/</i> .		LETTER A → ¢ x0D chinook pipa letter I · Circle vowel 2E ℃ 0B 〇 CHINOOK PIPA LETTER OO 2F
9 O A 0	\bigcirc			 Circle vowel C O CHINOOK PIPA LETTER OW Circle vowel OD C CHINOOK PIPA Circle vowel CHINOOK PIPA CHINOOK PIPA CHINOOK PIPA CHINOOK PIPA CHINOOK PIPA
в О с О	\bigcirc	+ *		$\begin{array}{c} 0D \\ LETTER I \\ \rightarrow \sim x2D \text{ chinook pipa} \\ letter E \\ 0E \\ CHINOOK PIPA \\ \end{array}$
D (Ŏ	v		0E CHINOOK PIPA LETTER U 34 ⊕ CHINOOK PIPA LIKALISTI SIGN Non-Conjoining Consonant 35 <reserved></reserved>
E C F •	+	Ċ		$0F \cdot CHINOOK PIPA$ LETTER H $36 $ $37 $ $38 $
Short Line Consonants 00 ¹ CHINOOK PIPA LETTER P · number 1 01 ⁻ CHINOOK PIPA LETTER T · number 2				Long Line Consonants 39 <reserved> 10 CHINOOK PIPA 3A <reserved> 11 CHINOOK PIPA 3C <reserved> 11 CHINOOK PIPA 3D <reserved></reserved></reserved></reserved></reserved>
02 \ CHINO · numbe 03 / CHINO · numbe · writt left	r 3 OK PIPA r 4		RK	12 \ CHINOOK PIPA LETTER V Punctuation 13 / CHINOOK PIPA LETTER G • written down and to the left Punctuation 3E = CHINOOK PIPA PUNCTUATION FULL STOP
04 [✓] CHINO · numbe	r 5	LETTER		14 / CHINOOK PIPA LETTER R • written up and to the rightConcatenating Control Character 3F3FCONCATENATOR • signifies abbreviations and initialisms

3 of 7

8/7/2009 10:46 AM

Modified Arc Consonants15 <reserved>16)CHINOOK PIPA LETTER NG17→CHINOOK PIPA LETTER CH18→CHINOOK PIPA LETTER TS</reserved>	 • shape shown is arbitrary and is not visibly rendered → 10A3F Kharoshthi Virama
W- Vowels 19 ♀ CHINOOK PIPA LETTER WO	
1A O CHINOOK PIPA LETTER WA	
1B 🗘 CHINOOK PIPA LETTER WOO	
1C 🖸 CHINOOK PIPA LETTER WOW	
1D 🔿 CHINOOK PIPA LETTER WE	
1E <reserved></reserved>	

Zero Width Non-Joiner and Zero Width Joiner

The Chinook Pipa script has a standard cursive conjoining behaviour effected by the use of the Zero Width Non-Joiner. The Zero Width Non-Joiner (ZWNJ) encodes a syllabic break, overriding the standard conjoining of all characters except H and X. The ZWNJ and Zero Width Joiner have no effect when adjacent an H or X. This contrasts slightly with its usage in other scripts, where it generally encodes for conjoining *forms* rather than conjoining *behaviour* in Chinook Pipa. Also, because of conjoining behaviour being the default, there is no general use for the Zero Width Joiner, except the abnormal instance of joining Chinook Pipa characters in running text to characters from other scripts with joining properties, the default behaviour in this instance being unjoined. For lists consisting of Chinook Pipa characters and, for example, a medial form of an Arabic character (eg. Sh + medial Alef + M) a sequence like Sh + ZWNJ + ZWJ + Alef + ZWJ + ZWNJ + M should be used.

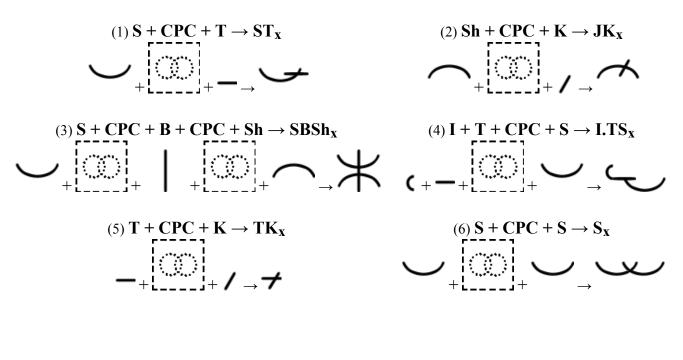
Cursive Conjoining

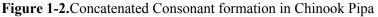
The most common form of character interaction is that of the cursive connection. The termination of the stroke of an initial character leads directly into the beginning of the next character. Circle vowels are connected on their perimeter at a tangent to adjacent consonants and I/E. The vowels "I", "E", and "U" rotate to connect without angle to a preceding character and with minimal angle into following characters. The letter "Oo" connects with preceding characters as normal, but in flowing handwritten text most often cursively connects with a following character at the nub, which orients to a position dictated by the adjacent consonants. Words are often split into nominally syllabic units with ZWNJ, H, or X, with breaks occuring between syllables containing circle and W- vowels and in other environments, such rules being generally algorithmic, but best handled manually or with spell-checking and other word-processing features, given variations in syllable breaking among the dozen or so Chinook Pipa orthographies for different languages.

Chinook Pipa Concatenator

Normally, Chinook Pipa letters conjoin together cursively. There is, however, a variant joining behaviour, in which adjacent consonants will overlap, signifying an abbreviation, initialism, or acronym (denoted CC_x). The Chinook Pipa Concatenator (CPC:

, U+x3F), signifies this alternate concatenating behaviour, much like the Virama in Indic scripts indicating conjunct letters. The CPC is interlocuted between the effected consonants, signifying the concatenating interaction of the two letters.







Combining diacritical marks on vowels

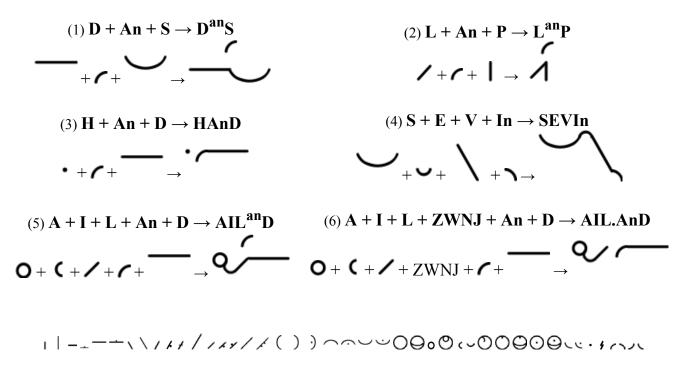
The Chinook Pipa script uses several combining diacritical marks, including an over-dot, underdot, diaeresis, and under diaeresis. The macron, under-macron, acute, and breve are also found in Salishan texts. These last four do not place directly above (or below) their base letter, but are instead shifted right, so their left-hand extreme is directly over the center of the base letter. The under macron has only been found in combination with acute, as some writers (mostly LeJeune) move the macron below a vowel to avoid collision with the acute placed above that vowel.

$$||_{-+} - - + \langle \langle , , , , , / \rangle \langle () \rangle \rangle = - - \langle \langle , \rangle \rangle \\ = - + \langle \langle , , , , , , \rangle \rangle$$

Nasal Vowels

The Chinook Pipa nasal vowels have a combining behaviour unlike any other characters. In certain circumstances, they take the form of a diacritic mark over the intersection of the two adjacent characters, and in others they will render inline, just as a regular letter. The nasal vowels will default render displaced - as a diacritic - if adjacent two conjoining consonants, ie not H, X, or separated by ZWNJ. In all other circumstances - adjacent an H, X, ZWNJ, or vowel - a nasal vowel will be rendered cursively connected to an adjacent (non-H/X) consonant or in isolation. ZWNJ will override displaced rendering by splitting an adjacent consonant into another sy llable and enabling in-line rendering attached to the other consonant. A displaced nasal vowel will render below the intersection of adjacent consonants if room is not available above.

Figure 1-5.Nasal Vowel rendering



Non-conjoining characters

The other characters in the Chinook Pipa - the letters "H" and "X", Full Stop, and "Likalisti" sign - do not typographically interact with other letters. The letters "H" and "X" split syllables fore and aft. The Chinook Pipa Full Stop character is used fairly frequently like a period or colon, probably due to these punctuations' similarity to Chinook Pipa letters. The logograph "Likalisti", meaning *eucharist* is found in several texts and should be spaced as a word, not as a syllable.

Vowel Orientation

Chinook letters conjoin regularly by fairly simple rules. All consonants have a stroke direction - for P/B, F/V, K/G, M/N, and all variants, the stroke direction is top-down; for T/D, L/R, Sh/S, and variants, stroke direction is left to right. Consonants join with the stroke termination of the first consonant marking the beginning of the second consonant's stroke. Consonants, I, and E join to circular vowels and circular vowels to consonants, I, and E at tangent angles. In source texts, the circles are actually continuations of the consonant strokes moving into and out of the circular vowel form. Vowels often combine beneath and to the right of consonants, but generally above for the pattern T/D/L/R preceding a circle vowel plus S/Sh/N/P/B/K/G, or the pattern L/R + circle vowel + T/D. Circle vowels usually combine inside the arc consonants, M/N/Sh/S/Ng/Ch/Ts. I and U almost exclusively follow the "in from the top or left, out down or right" rule, and E orients exactly opposite when joining a single letter, except K/G, L/R and variants. An isolated E is also known to render upside-down, such a distinction necessitating markup either outside the scope of the Unicode Standard or through a variation selector. At this time, it is thought best to leave this distinction outside Unicode until it can determined whether such distinction is contrastive. These rules having been given, the down/right rule will always be intelligible, though less elegant than contextual implementations.

Archives of the Kamloops Wawa 1891-1900 (subscription required)

Dictionary of the Chinook Jargon, by George Gibbs, Echo Library ISBN 1-40680-924-1

Chinook:.... A History and Dictionary, by Edward Harper Thomas, 1935, Metropolitan Press, Portland, OR)

Documentation can be found <u>here.</u>