

Suggestions for Changes to UCAS

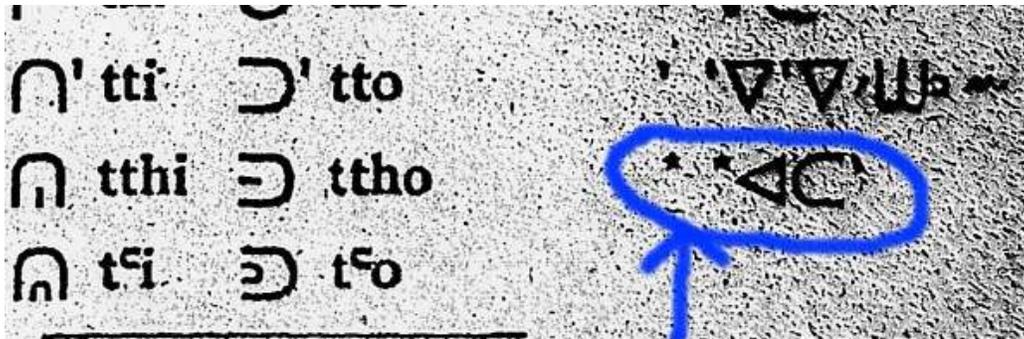
I would like to suggest to the Unicode community the following observations relating to the Unified Canadian Aboriginal Syllabics range. My goal (see www.languagegeek.com) is to enable all of the North American languages to be properly and accurately written on the Internet, and computers in general. Here I will focus specifically on the languages which are currently using or historically used (and still are in some communities) syllabics.

Some conventions are used below. All Unicode character names are in majuscule, and “Canadian Syllabics” has been abbreviated to CS. Hexadecimal Unicode indices are in parentheses and prefixed with “U+”. All sources cited are linked to the languagegeek.com bibliography. A “final” is the Syllabics term for a character which represents a consonant only, not a consonant + vowel, so CS FINAL GRAVE (U+1420), CS CARRIER H (U+144B) and CS NASKAPI SKW (U+150A) would all be examples of “finals”. I use the term “syllabic” to refer to a consonant + vowel character. A series is a row of characters on a syllabic chart, so in Misnamed Characters Note 1, “tta, tte, tti, tto” would be the tt-series.

Characters that may merit discontinuation

1. The asterisk * character (U+156F) appears on the code–page chart as **, and is named CS TTH. This is a misreading of the syllabarium chart used by the French Missionaries for Chipewyan—probably from the 1904 publication *Prières Catéchisme et Cantiques en langue Montagnaise ou Chipeweyan*. The chart in this book has been reprinted in most if not all “scripts of the world” type books. The relevant excerpt from the chart is shown below.

Unlike most other syllabics charts, this one does not have a column of finals to the right of the consonant–vowel syllabics. Instead, it simply has a list of all the finals, which do not correspond with the syllabics series on the same row. Thus, the CS WEST–CREE P (U+144A) ' final which appears to the right of the “tta” row is not the sound “tt”, but is instead “h”. The blue circled asterisk is not “tth”,



but is in fact a symbol which indicates a proper name, in this case *ᑎ` /*ada/ (Adam). A second glitch on the Unicode code–page chart is that this character is written with two asterisks “**”, when in fact on the chart above, the first asterisk is the character itself, and the second is part of the example. I believe this should definitively be fixed.

2. Please see below “Questions”, note 3.

Missing Characters

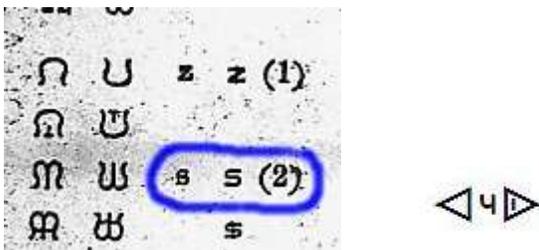
Blackfoot

1. In Blackfoot, a raised “equals sign” = is used much as the “CS FINAL MIDDLE DOT” (U+1427) is in Cree: to indicate a /w/ between the consonant and vowel of the syllabic. In the graphic below, the red arrow is pointing to this character, which in combination with the “CS BLACKFOOT KA” (U+15BD) before it, gives the sound /kwa/ ᑭᑦ. This character is vital to writing Blackfoot, should be added. This example taken from *Stockten 1888*.

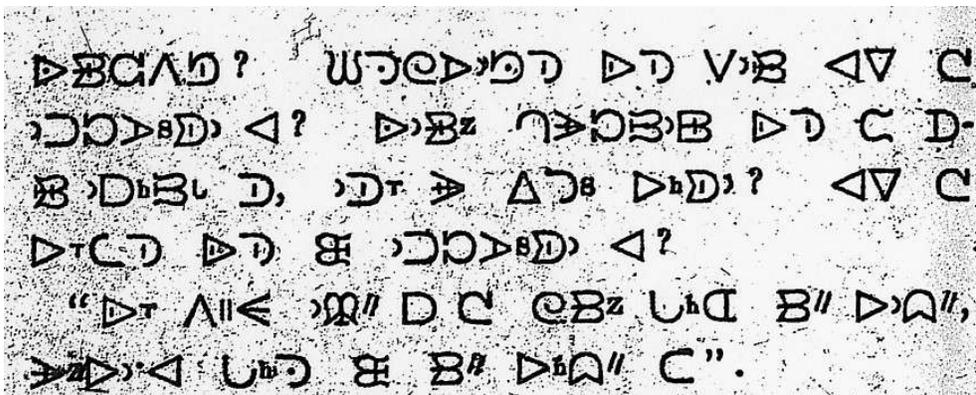


Carrier Dene

1. A few finals are missing from Unicode which are used in Carrier. Information for Carrier is from *Poser 2000*. There is an important graphical distinction between the finals used for /s/ and /s̥/. The former is a small serif “s” written mid-line, while the latter is a small sans-serif “s” written mid-line. This is exemplified below by the blue circle (from *Morice 1894*). Unicode lists only one version (U+1506) CS ATHAPASCAN S. A second character, an upside-down mid-line small “h” ʘ is used for load words with /f/ or /v/ sounds. The second example below is /ave/. These two finals should be added.



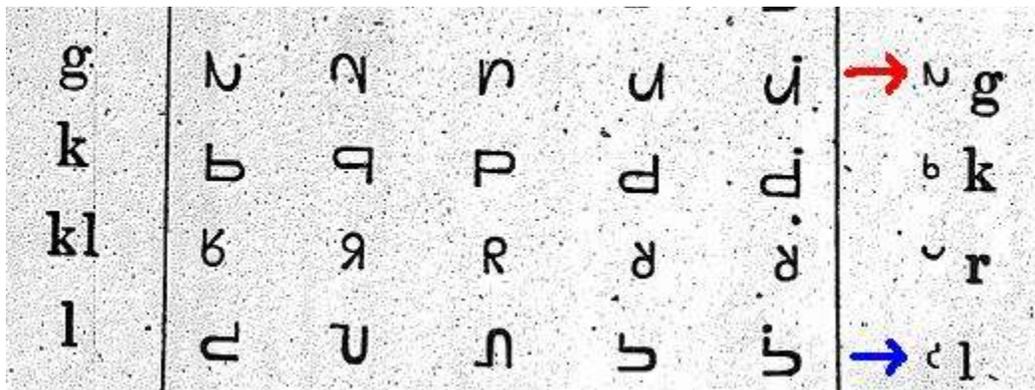
2. In examples of Carrier, the finals are virtually always mid-line. This is purely stylistic, but see Dene note 2 about how final placement is phonetically important. This example is from *Morice 1894*, note that all of the finals are mid-line, not top-line as in Unicode.



Please see Note 1 under "Questions" below for discussion of the Carrier glottal stop.

Dene (Chipewyan, Slavey, Hare, Beaver)

1. In early Chipewyan texts (the example below is from [Kirkby 187.](#)), a-finals—like those used in Eastern dialects of Cree—were used instead of the western-finals which were employed in all later Chipewyan texts. For the most part, Unicode includes these finals under names such as CS K (U+1483), however, some of the Chipewyan series are not to be found in Cree, and are thus absent. The two missing finals are shown in the graphic below. The blue arrow indicates a raised small version of CS WEST-CREE LA (U+154D) ᑭ. The red arrow points to a raised /ga/ syllabic ᑭ. The g-series here corresponds to CS SAYISI HE, HI, HO, HA (U+15C0–15C3). Though no longer used today, for historical purposes, these two characters may be added.

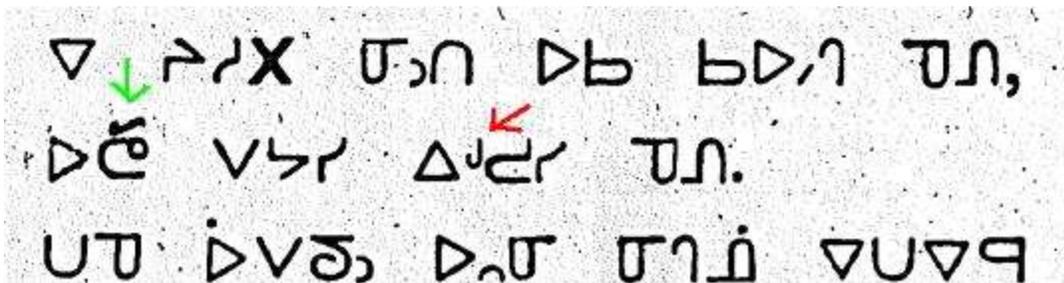


2. In Dene texts, it is vitally important to distinguish between the position of the finals, top-line, mid-line, or bottom-line. The same shape character may have a completely different value depending on how high up from the baseline it is. In the first example below (from [LeGoff 1890](#)), the red arrow shows the top-line Chipewyan character which resembles CS WEST-CREE M (U+14BC) ᑭ. The blue arrow points to a mid-line version of the same shape. The top-line is a diacritic which indicates that the following sound (the mid-line “s”) is pronounced /t/ instead of /l/ if the top-line diacritic were missing. The mid-line version has the sound /m/. The word ᑭᑭᑭᑭ would be pronounced along the lines of /yałtiy/. Here the character ᑭᑭ is the historical equivalent of ᑭᑭ.

The English Anglican style of writing Dene is different in some important ways. With regards to final placement, where the French system distinguishes top- and mid-line finals, the English differentiate between top-, bottom-, and a few mid-line finals. In the first example below, from a South Slavey text ([Reeve 1900](#)), the red arrow shows a top-line character like CS C (U+14A1) ᓅ, whereas the blue indicates the base-line variety. When at the top, finals in the Anglican tradition are diacritics modifying the following sound (much like the Catholic examples above). So, a top-line CS C is combining with the following CS WEST-CREE LE (U+1544) to produce /tli/ ᓅᓴ (where /tli/ is a single phoneme). The base-line final ᓅ is the /t/ phoneme alone.



This system holds true with other finals as well. The base-line version of CS SAYISI TH (14A2) ᓵ is a final pronounced /th/ or /dh/. The next example ([Kirkby 1881c](#)), has a top-line CS SAYISI TH ᓵ, which combines with the following syllabic CS WEST-CREE LA (U+154D); together ᓵᓴ they sound like /ʔa/. If the top-line CS SAYISI TH were written at the baseline ᓵᓴ, the sound would be /thla/ or /dhla/. Moreover, if in the above example, the baseline CS SAYISI TH were at the top-line, instead of /thli/, the sound would be /ti/.



These examples should make it clear that the Dene Syllabics system distinguishes top-line from mid-line and baseline finals in very important ways. I have not

included all of the characters which can change position to make different phonemes (but I could if necessary).

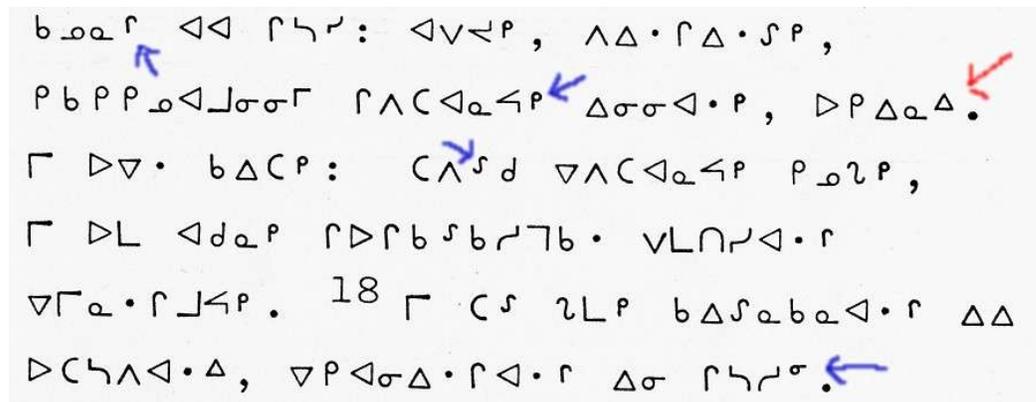
As for how to render this in Unicode, I see two possibilities.

a) Encode three unique characters for each final, i) top-line, ii) mid-line, iii) baseline. First, this solution is not terribly encoding-efficient. Also, it sets up different encodings for glyph variants of different styles. For example, some Cree fonts place the CS FINAL RING (U+1424) ° at the top-line, others at mid-line. This differentiation should be based on the font, not on Unicode encoding. An opentype substitution would be a much better solution for stylistic differences I think.

b) Add two or three non-spacing “characters” to the UCAS range, which tell the software where to put the final: (top), mid, or base. This also allows much more flexibility for writers who want control over how high up their finals reside. See Note 4 under “Questions” for whether a top-final would be required.

Ojibway

1. Several dialects of Ojibway in Northern Ontario have a unique means of writing finals. It follows the logic of the Eastern finals (a small top-line version of the a-syllabic), but instead it employs the i-version of the syllabic. The *example* below, points out several i-series finals.

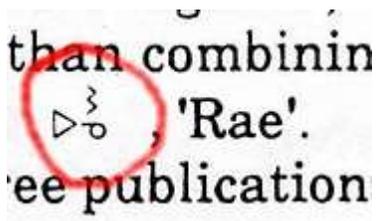


The first blue arrow is the i-final version of CS C (U+14A1) °, the second is of

CS K (U+1483) ᐅ, the third is of CS SH (U+1525) ᓄ, and the last is of CS N (U+14D0) ᓂ. The red arrow points out the CS GLOTTAL STOP (U+141E) ᐃ which already exists in Unicode. The full list of i-finals is below. The top row contains the i-finals, the bottom row the a-final equivalents. The entire series of i-finals should be added to Unicode.

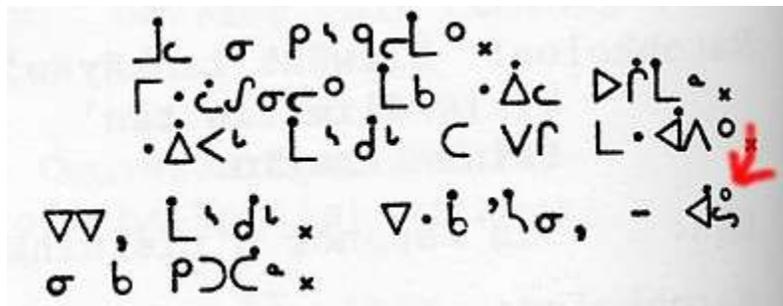
ᐱ	ᐃ	ᐅ	ᐇ	ᐉ	ᓂ	ᓄ	ᓆ	ᐃ	ᐅ
<	c	b	ᓄ	L	a	ᓄ	ᓆ	o	ᓄ

2. Some dialects of Ojibway use the raised “i” and “r” finals (U+14EB ᓄ and U+1551 ᓄ) as distinct characters, while others place these above an n-series syllabic as a diacritic. Two non-spacing characters would be required to accommodate this.



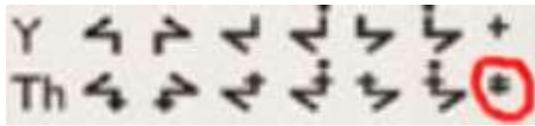
Cree

1. The UCAS code-chart incorrectly labels U+141E ᐃ as a Moose Cree (Y). In fact, the Moose Cree y-final is a small ring diacritic located above the syllabic character. The Inuktitut characters like CS AAI (U+1402) ᐃ and CS PAAI ᐃ (U+1430) would be read /iy/ and /piy/ in Moose Cree. The following graphic, from *Ellis 1983*, displays this diacritic clearly atop CS SHA (U+1515), marked by the red arrow, in the word ᐃᓄ /aashay/ “now”.

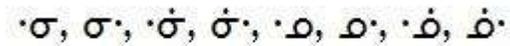


Unicode would have to encode a large number of pre-composed glyphs of a Moose Cree syllabic with the y-ring on top—not to mention those in combination with CS FINAL MIDDLE DOT (U+1427). Instead, using the ring non-spacing diacritic would be preferable. But U+141E is not a Moose Cree /y/. Some speakers prefer to place the small ring not above the syllabic character, but to the right like other finals, thus /aashay/ is $\langle \omega^\circ >$. This small ring is absent from Unicode, and should be included. This small ring should not be confused with the larger CS FINAL RING (U+1424) °, which in the example above appears just to the left of the tail of the arrow.

2. The Woods Cree dialect (labelled by Unicode as TH-Cree, U+15A7–U+15AE) does not use an Eastern-style, raised a-final, as shown in U+15AE ʔ. To my knowledge, U+15AE would not be used by anyone. The final I have seen used (from the La Ronge Cree community) is missing from Unicode. The red circle indicates the final *. This character is required to write this dialect of Cree.



3. I am curious to know why the n-series of Cree syllabics (U+14C0–14D2) is missing half of the w-dot characters, namely: nwi, nwii, nwo, nwoo (both eastern and western versions).



4. For Y-Cree, please see note 1 under "Questions" below.

All languages:

The correct form of the hyphen in Syllabics is a shortened equals sign, the example below is from the same Ojibway text mentioned above. Should this hyphen get its own Unicode encoding? A regular equals sign doesn't necessarily look right, and will not wrap properly. A single dash hyphen is unacceptable because it would conflict with CS FINAL SHORT HORIZONTAL STROKE (U+1428) -.



w	·▽	·△	·▷	·◁	◦	◦
y	◁	▷	▽	△	◦	◦
h						

2. Would anybody know the sources used for the characters referenced as “Sayisi” (from Tadoule Lake, Manitoba)? I am completely unaware of certain characters, such as U+14BE, U+14BF, U+1541. In general, the Sayisi characters match the English–tradition Dene syllabics (as one would expect), but I have never come across the three characters above.

3. Why are CS FWAA (U+155A–B) ·◁◦ ◁◦, CS THWAA (U+1567–8) ·◦◦ ◦◦, and CS RWAA (U+154E–154F) ·◦◦ ◦◦ part of UCAS? What about “fwe, fwi, fwii, fwo, fwoo, fwa”, “thwe...”, and “rwe...” (both eastern and western versions)? Why do the long ā glyphs merit inclusion? I would guess that someone was reading directly off a syllabics chart, not realising that the 6 glyphs above were just examples of entire series. The CS F..., CS TH..., and CS R..., are for load words only in Cree. A **combing diacritical mark “mid-dot”** would be useful for those syllabics which were not encoded as composed characters: e.g. “fwe”. If one were to use the CS FINAL MIDDLE DOT (U+1427) plus the syllabic character (in this case, CS FE [U+1553]), it would not space properly. This combining mid dot would also accommodate the missing missing half of the nw–series (Cree Note 3). There would in fact need to be two cdm mid-dots: one for Eastern dialects, one for Western dialects.

4. In some Dene systems, super script F, V, r, and l (F, V, r, l) are used as finals to indicate these sounds from European languages. Carrier Dene also uses a regular

