1. Introduction. The Wancho language has some 51,000 speakers and is used chiefly in the southeast of Arunachal Pradesh, as well as in Assam and Nagaland. There are 64 villages in Longding, 4 villages in the Tirap district, 128 villages in Nagaland, over 23 villages in Myanmar (where it may be called Chuyo), 3 in Bhutan and 4 in Assam, all where the language is spoken. Wancho is a Sino-Tibetan language of the Konyak or Northern Naga subfamily of the Sal family spoken in eastern India, parts of Bangladesh, and Myanmar. There are three main dialect areas, Northern, Central, and Southern. The script was devised between 2001 and 2012 by Banwang Losu, a teacher at a government middle school in his home village in Longding District, Arunachal Pradesh. The Wancho script is taught today in schools.

2. Processing. Wancho is a simple alphabetic script comprised of letters which represent both consonants and vowels. There are no conjunct characters. Diacritical marks are used on vowel letters to indicate tone. A few characters are similar to other characters, distinguished by a diacritical mark, but this dot is not productive; while five of them appear in pairs in alphabetical order (𱅠 AA a, 𱅡 A ə; 𱅢 SA s, 𱅣 SHA ś; 𱅤 JA j, 𱅥 ZA z; 𱅦 O o, 𱅧 AU au; 𱅨 AANG aŋ, 𱅩 ANG āŋ), the dotted forms of both 𱅪 RA r and 𱅫 UEN u (𱅫 TRA ตร̥ and 𱅫 UEN ɯ̃) are sorted at some distance from them.

3. Orthography. Tones are marked in various ways in different varieties of Wancho. An absolute system of tone marking would not work for Wancho because, for instance, in an Upper Wancho dialect, it is the low tone which is the most common tone, but in a Lower Wancho dialect, it is the high tone which is the most common. The convention is to leave the most common tone unmarked, to mark the second-most common tone with one diacritical mark, and to mark the least-common tone with a double diacritical mark. Different diacritical marks are used to distinguish language varieties whose unmarked tone differs. Syllables checked with the glottal stop are neutral as to tone, and are also unmarked.

3.1 Tone marks not present in N4785. The Wancho script was developed for all the Wancho languages, but Losu 2013 focussed on Northern Wancho. In the orthography used or that variety, the most common low tone was marked with a single vowel (as in 𱅡 gì ‘cane’), the mid tone with two vowels (as in 𱅡 gì ‘thread’), and the least-frequent high with three vowels (as in 𱅡 gì ‘incise’). It became clear that this mechanism was cumbersome, and when the description of Southern Wancho (Burling and Wangsu 1998) was examined, it became clear that tone distribution (in terms of frequency) was not uniform and that this scheme would prove so confusing to users that they were likely to avoid doubling and tripling vowels altogether. The four tone marks proposed here were devised in consultation with Bangwang Losu, in two pairs, one to be used with Northern Wancho and one with Southern. In a sense
this is analogous to the use of the acute accent in Irish and the grave accent in Scottish Gaelic: the

different diacritics help a reader identify the language and orthography at a glance. Other Wancho
languages are not well described (and it could take decades before that is accomplished). See Figure 11.

In an Upper Wancho dialect (tones are low, mid, high, checked):

\[\text{̣̣} \text{gi} \text{ ‘cane’} \quad \text{̣̣} \text{gi} \text{ ‘thread’} \quad \text{̣̣} \text{gi} \text{ ‘incise’} \quad \text{̣̣} \text{gi} \text{ ‘wilting’} \]

\[\text{̣̣} \text{tū} \text{ ‘blunt’} \quad \text{̣̣} \text{tū} \text{ ‘iron’} \quad \text{̣̣} \text{tū} \text{ ‘sec’} \quad \text{̣̣} \text{tū} \text{ ‘thorn’} \]

\[\text{̣̣} \text{kā} \text{ ‘legs apart’} \quad \text{̣̣} \text{kā} \text{ ‘go’} \quad \text{̣̣} \text{kā} \text{ ‘door’} \quad \text{̣̣} \text{kā} \text{ ‘soil’} \]

\[\text{̣̣} \text{kè} \text{ ‘buttock’} \quad \text{̣̣} \text{kè} \text{ ‘weak’} \quad \text{̣̣} \text{kè} \text{ ‘—’} \quad \text{̣̣} \text{kè} \text{ ‘—’} \]

\[\text{̣̣} \text{ki} \text{ ‘quarrel’} \quad \text{̣̣} \text{ki} \text{ ‘dog’} \quad \text{̣̣} \text{ki} \text{ ‘hatch’} \quad \text{̣̣} \text{ki} \text{ ‘to salve’} \]

\[\text{̣̣} \text{ò} \text{ ‘noise’} \quad \text{̣̣} \text{o} \text{ ‘bird’} \quad \text{̣̣} \text{o} \text{ ‘protect’} \quad \text{̣̣} \text{o} \text{ ‘—’} \]

In a Lower Wancho dialect (tones are high⁴, falling⁵², low¹¹, checked):

\[\text{̣̣} \text{zi} \text{ ‘thread’} \quad \text{̣̣} \text{li} \text{ ‘water buffalø’} \quad \text{̣̣} \text{zi} \text{ ‘die’} \quad \text{̣̣} \text{lì} \text{ ‘warm’} \]

\[\text{̣̣} \text{̣意识形态} \text{ ‘braid’} \quad \text{̣意识形态} \text{ ‘hair of head’} \quad \text{̣意识形态} \text{ ‘door’} \quad \text{̣意识形态} \text{ ‘penis’} \]

\[\text{̣意识形态} \text{ ‘dog’} \quad \text{̣意识形态} \text{ ‘—’} \quad \text{̣意识形态} \text{ ‘point’} \quad \text{̣意识形态} \text{ ‘small’} \]

The diphthongs /ai/ and /au/ may be short or long; /au/ is spelt somewhat differently when long. These

are spelt /ai/ and /au/ here, though another transcription might be /ai/ and /ai/. The tone mark is

placed on the first vowel in a diphthong.

\[\text{̣意识形态} \text{ ‘cough’} \quad \text{̣意识形态} \text{ ‘sputum’} \quad \text{̣意识形态} \text{ ‘kick’} \quad \text{̣意识形态} \text{ ‘beat’} \]

\[\text{̣意识形态} \text{ ‘give, let’s’} \quad \text{̣意识形态} \text{ ‘bent’} \quad \text{̣意识形态} \text{ ‘plant’} \quad \text{̣意识形态} \text{ ‘flick’} \]

\[\text{̣意识形态} \text{ ‘half’} \quad \text{̣意识形态} \text{ ‘—’} \quad \text{̣意识形态} \text{ ‘—’} \quad \text{̣意识形态} \text{ ‘—’} \]

\[\text{̣意识形态} \text{ ‘he, she’} \quad \text{̣意识形态} \text{ ‘step’} \quad \text{̣意识形态} \text{ ‘run-off’} \quad \text{̣意识形态} \text{ ‘give’} \]

3.1 Vowels not present in N4785. The other difference between the repertoire concerns the addition of

two short vowels, ̣ a and ̣ ə, which are ̣ a and ̣ ə augmented with the use of the dot diacritic.

Minimal pairs exist: ̣意识形态  ‘look upwards’, ̣意识形态  ‘banana leaf’ (tones unknown at present for this

pair); ̣意识形态  ‘sky’, ̣意识形态  ‘shoulder’. The previous orthography was defective and could not

support these distinctions.

4. Non-alphabetic characters are used in Wancho. Ordinary generic punctuation marks like comma and

full stop and question mark are used in Wancho and can be represented by existing UCS characters.

5. Digits. U+1E2F0..U+1E2F9 are the decimal digits 0–9. Common punctuation and operators are used

for mathematical operations. See Figure 3.

6. Ordering. The ordering of the character set follows that of the code chart. The first four characters are

given in an order which resembles the Aramaic order used by many alphabets (a b c g d, sort of a

hybrid of Latin and Greek ABCD) but this seems to be incidental and there is no particular similarity to

that or any other order otherwise, apart from a division between sounds found in Hindi (AA to U) and

sounds unique to Wancho (LLHA to YIH). Dotted letters are sorted as separate from their undotted

counterparts.

7. Character names. The character names here use the usual UCS transcription for the scripts of India:

̣, ̣ a, ̣ a, ̣ ba, ̣ ca, ̣ tf, ̣ da, ̣ ga, ̣ ya, ̣ ha, ̣ la, ̣ ma, ̣ na, ̣ pa, ̣ ta, ̣ tha, ̣ fa, ̣ sa, ̣ sha, ̣ ja, ̣ za, ̣ wa, ̣ va, ̣ ka, ̣ o, ̣ au, ̣ ra, ̣ ma, ̣ kha, ̣ ha, ̣
8. Unicode Character Properties.

9. References

10. Acknowledgements. This project was made possible in part by a grant from the U.S. National Endowment for the Humanities, which funded the Universal Scripts Project (part of the Script Encoding Initiative at UC Berkeley) in respect of the Wancho encoding. Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment for the Humanities.
Figures.

**Figure 1.** Chart of the Wancho alphabet. The gap appears between the letters which are used for sounds common to other languages of India, and sounds which are considered unique to Wancho. From Losu 2013:31. Since the publication of Losu 2013 analysis of the use of Wancho script for multiple varieties of Wancho has led to the addition of the following characters, which are part of this proposal:
Figure 2. Some minimal pairs in Wancho. From Losu 2013:86–87. On the left the words are given in an earlier orthography which marks tone by doubling and tripling vowels and nasals. On the right is the reformed orthography, which has a vowel ꤪ for short a and the second and third tones (in the case of this variety of Northern Wancho, mid and high tones).
Figure 3. Examples of arithmetical operations in Wancho. From Losu 2013:96–97.

Figure 4. Example of the Wancho ngun sign, used to indicate rupees. From Losu 2013:119.
Figure 5. Examples of words beginning the alphabet in Wancho script, Latin script, Devanagari script, with English and Hindi translations. Variant readings like �ระหว่าง���=./v chahnu [tʃaʔnù] / 𜙞𜖭 thahnu [θaʔnù] indicate dialect variation. From Losu 2013:44–45.
**Figure 6.** A Wancho poem. From Losu 2013:124.
Figure 7. A Wancho song with transliteration. From Losu 2013:125.
Figure 8. A page from one of the Middle School examinations in Wancho.
**Figure 9.** Students taking an examination in Wancho script.

**Figure 10.** Wancho keyboard layout.
To,
The JTC1/SC2/WG2 and UTC

Subject: Wancho Orthographies

Sir/Madam,

With profound respect, the undersigned is requested to encode the following Combining Marks as shown in the Proposal of the Wancho script in the UCS, as they are an important part of our official Orthography. Without those four characters the whole sets of Orthography is incomplete.

The four characters are below:

1.  
2.  
3.  
4.  

Single dot is called Tup, double dot is called Tupni. The triple dot is Tupram and if dots are more than three then it is read as Tup Mang. Only one and two are used in Wancho orthography.

Similarly, single feather is called Koi (derived from Okoi “feather”), double Koi is called Koini. Three feathers, is called Koiram and more than three, it reads as Koi Mang. Only one and two are used in Wancho orthography.

This is for your kind information and necessary action please.

With Great Regards

Banwang Losu  
Author of Wancho Script  
& Executive Director  
Wancho Literary Mission

Figure 11. Letter from Banwang Losu affirming the use and names of the four diacritical marks.
Figure 12. Relative positions of the four diacritical marks over the vowels of Wancho. Generally they are centred over the chief curve at the top of the character.
A. Administrative

1. Title
Proposal to encode the Wancho script in the UCS

2. Requester’s name
UC Berkeley Script Encoding Initiative (Universal Scripts Project)

3. Requester type (Member body/Liaison/Individual contribution)
Liaison contribution.

4. Submission date
2017-10-22

5. Requester’s reference (if applicable)

6. Choose one of the following:
6a. This is a complete proposal
Yes.

6b. More information will be provided later
No.

B. Technical – General

1. Choose one of the following:
1a. This proposal is for a new script (set of characters)
Yes.

1b. Proposed name of script
Wancho.

1c. The proposal is for addition of character(s) to an existing block
No.

1d. Name of the existing block

2. Number of characters in proposal
59.

3. Proposed category (A-Contemporary; B.1-Specialized (small collection); B.2-Specialized (large collection); C-Major extinct; D-Attested extinct; E-Minor extinct; F-Archaic Hieroglyphic or Ideographic; G-Obscure or questionable usage symbols)
Category A.

4a. Is a repertoire including character names provided?
Yes.

4b. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document?
Yes.

4c. Are the character shapes attached in a legible form suitable for review?
Yes.

5a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?
Michael Everson via Banwang Losu

5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:
Michael Everson, Fontographer.

6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?
Yes.

6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?
Yes.

7. Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?
Yes.

8. Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script.
See above.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.
Yes. N-4785.

2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?
Yes.

2b. If YES, with whom?
Banwang Losu, Stephen Morey, Andrew Cunningham, Yhilan Ko

2c. If YES, available relevant documents

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?
Contemporary use by Wancho speakers.

4a. The context of use for the proposed characters (type of use: common or rare)
Common.

4b. Reference

5a. Are the proposed characters in current use by the user community?
Yes.
5b. If YES, where?
**In Arunachal Pradesh.**

6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?
No.

6b. If YES, is a rationale provided?

6c. If YES, reference

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?
**Yes.**

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?
No.

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?
No.

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?
No.

10b. If YES, is a rationale for its inclusion provided?

10c. If YES, reference

11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?
No.

11b. If YES, is a rationale for such use provided?

11c. If YES, reference

11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
No.

11e. If YES, reference

12a. Does the proposal contain characters with any special properties such as control function or similar semantics?
No.

12b. If YES, describe in detail (include attachment if necessary)

13a. Does the proposal contain any Ideographic compatibility character(s)?
No.

13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?