

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation internationale de normalisation
Международная организация по стандартизации

Doc Type: Working Group Document

Title: Revised proposal for encoding the Ol Chiki script in the UCS

Source: Michael Everson, R. C. Hansdah & N. C. Murmu

Status: Expert Contribution

Action: For consideration by JTC1/SC2/WG2 and UTC

Date: 2002-11-03

This is a revision of N1956, and contains the proposal summary form.

A. Administrative

1. Title

Revised proposal for encoding the Ol Chiki script in the UCS.

2. Requester's name

Michael Everson, R. C. Hansdah & N. C. Murmu

3. Requester type

Expert contribution.

4. Submission date

2002-11-03

5. Requester's reference

6a. Completion

This is a complete proposal.

6b. More information to be provided?

No.

B. Technical -- General

1a. New script? Name?

Yes. Ol Chiki, also known as Santali, Ol, Ol Cemet' and Ol Ciki.

1b. Addition of characters to existing block? Name?

No.

2. Number of characters

48

3. Proposed category

Category A.

4. Proposed level of implementation and rationale

Level 2 as it uses modifying diacritics in the Brahmic style.

5a. Character names included in proposal?

Yes. There are some issues yet to be resolved regarding the best transliteration for the character names.

5b. Character names in accordance with guidelines?

Yes.

5c. Character shapes reviewable?

Yes (see below).

6a. Who will provide computerized font?

James Kass via Michael Everson.

6b. Font currently available?

Yes.

6c. Font format?

TrueType.

7a. Are references (to other character sets, dictionaries, descriptive texts, etc.) provided?

Yes, see bibliography below.

7b. Are published examples (such as samples from newspapers, magazines, or other sources) of use of proposed characters attached?

Yes, attached.

8. Does the proposal address other aspects of character data processing?

Yes, see Unicode properties below.

C. Technical -- Justification

1. Contact with the user community?

No.

2. Information on the user community?

Speakers of the Santali language, whose population is 5,800,000, with 25%–50% literacy, according to the SIL *Ethnologue*.

3a. The context of use for the proposed characters?

To write the Santali language. Latin, Devanagari, Bengali, and Oriya scripts have also been used to write Santali.

3b. Reference

See bibliography.

4a. Proposed characters in current use?

Yes.

4b. Where?

In primary and adult education (general use).

5a. Characters should be encoded entirely in BMP?

Yes.

5b. Rationale

Contemporary use and accordance with the Roadmap.

6. Should characters be kept in a continuous range?

Yes.

7a. Can the characters be considered a presentation form of an existing character or character sequence?

No.

7b. Where?

7c. Reference

8a. Can any of the characters be considered to be similar (in appearance or function) to an existing character?

No.

8b. Where?

8c. Reference

9a. Combining characters or use of composite sequences included?

Yes.

9b. List of composite sequences and their corresponding glyph images provided?

Yes. There are some limitations on which ones can combine with which base characters (see below).

10. Characters with any special properties such as control function, etc. included?

No.

Naming issues

Characters are arranged in a 5 by 6 matrix, named in a conventional way as shown in the names list. The first characters in each row (LO, LA, LI, LU, LE, LOO) are vowels. Given here is a UCS name based on traditional transliteration, transliteration according to Zide 1996, and the transliteration from the ALA Romanization Handbook, and the transliteration proposed by Hansdah and Murmu in the Proposal part of this document.

Table with 5 rows and 6 columns of characters and their transliterations. Row 1: LO, LA, LI, LU, LE, LOO. Row 2: OT, AK, OKH, ONG, OL. Row 3: AT, AK, ACH, AM, AW. Row 4: AT, IS, IH, INY, IR. Row 5: UC, UTH, UNN, UY. Row 6: EP, EDD, EN, ERR. Row 7: OTT, OOPH, OOWN, OOH.

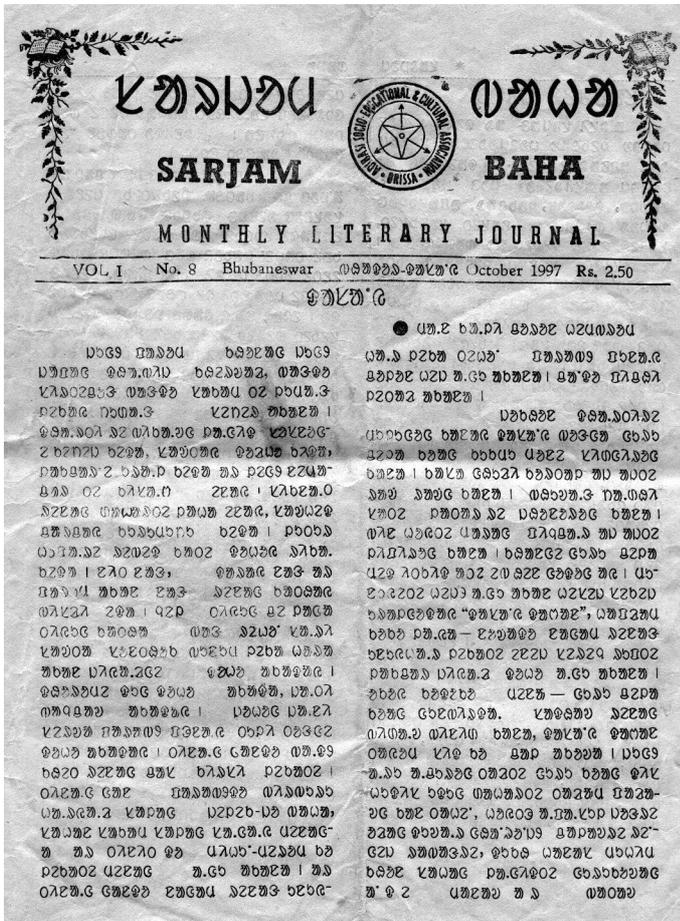


TABLE XXX - Row xx: OL CHIKI

	xx0	xx1	xx2
0	ᱠ	ᱡ	ᱢ
1	ᱣ	ᱤ	ᱥ
2	ᱦ	ᱧ	ᱨ
3	ᱩ	ᱪ	ᱫ
4	ᱬ	ᱭ	ᱮ
5	ᱯ	ᱰ	ᱱ
6	ᱲ	ᱳ	ᱴ
7	ᱵ	ᱶ	ᱷ
8	ᱸ	ᱹ	ᱺ
9	ᱻ	ᱼ	ᱽ
A	᱾	᱿	᱀
B	᱁	᱂	᱃
C	᱄	᱅	᱆
D	᱇	᱈	᱉
E	᱊	᱋	᱌
F	ᱍ	ᱎ	ᱏ

G = 00
P = 00

TABLE XXX - Row xx: OL CHIKI

dec	hex	Name
	00	OL CHIKI LETTER LA
	01	OL CHIKI LETTER AT
	02	OL CHIKI LETTER AG
	03	OL CHIKI LETTER ANG
	04	OL CHIKI LETTER AL
	05	OL CHIKI LETTER LAA
	06	OL CHIKI LETTER AAK
	07	OL CHIKI LETTER AAJ
	08	OL CHIKI LETTER AAM
	09	OL CHIKI LETTER AAW
	0A	OL CHIKI LETTER LI
	0B	OL CHIKI LETTER IS
	0C	OL CHIKI LETTER IH
	0D	OL CHIKI LETTER INY
	0E	OL CHIKI LETTER IR
	0F	OL CHIKI LETTER LU
	10	OL CHIKI LETTER UCH
	11	OL CHIKI LETTER UD
	12	OL CHIKI LETTER UNN
	13	OL CHIKI LETTER UY
	14	OL CHIKI LETTER LE
	15	OL CHIKI LETTER EP
	16	OL CHIKI LETTER EDD
	17	OL CHIKI LETTER EN
	18	OL CHIKI LETTER ERR
	19	OL CHIKI LETTER LO
	1A	OL CHIKI LETTER OTT
	1B	OL CHIKI LETTER OB
	1C	OL CHIKI LETTER OV
	1D	OL CHIKI LETTER OH
	1E	OL CHIKI PUNCTUATION MUCHAD
	1F	OL CHIKI PUNCTUATION DOUBLE MUCHAD
	20	OL CHIKI DIGIT ZERO
	21	OL CHIKI DIGIT ONE
	22	OL CHIKI DIGIT TWO
	23	OL CHIKI DIGIT THREE
	24	OL CHIKI DIGIT FOUR
	25	OL CHIKI DIGIT FIVE
	26	OL CHIKI DIGIT SIX
	27	OL CHIKI DIGIT SEVEN
	28	OL CHIKI DIGIT EIGHT
	29	OL CHIKI DIGIT NINE
	2A	OL CHIKI MU TUDAG
	2B	OL CHIKI GAHLA TUDAG
	2C	OL CHIKI MU-GAHLA TUDAG
	2D	OL CHIKI RELA
	2E	OL CHIKI PHARKA
	2F	OL CHIKI AHAD

Group 00

dec	hex	Name
		OL CHIKI LETTER LO
		OL CHIKI LETTER OT
		OL CHIKI LETTER OKH
		OL CHIKI LETTER ONG
		OL CHIKI LETTER OL
		OL CHIKI LETTER LA
		OL CHIKI LETTER AK
		OL CHIKI LETTER ACH
		OL CHIKI LETTER AM
		OL CHIKI LETTER AW
		OL CHIKI LETTER LI
		OL CHIKI LETTER IS
		OL CHIKI LETTER IH
		OL CHIKI LETTER INY
		OL CHIKI LETTER IR
		OL CHIKI LETTER LU
		OL CHIKI LETTER UC
		OL CHIKI LETTER UTH
		OL CHIKI LETTER UNN
		OL CHIKI LETTER UY
		OL CHIKI LETTER LE
		OL CHIKI LETTER EP
		OL CHIKI LETTER EDD
		OL CHIKI LETTER EN
		OL CHIKI LETTER ERR
		OL CHIKI LETTER LOO
		OL CHIKI LETTER OOTT
		OL CHIKI LETTER OOPH
		OL CHIKI LETTER OOWN
		OL CHIKI LETTER OOH
		OL CHIKI SINGLE PUNCTUATION
		OL CHIKI DOUBLE PUNCTUATION
		OL CHIKI DIGIT ZERO
		OL CHIKI DIGIT ONE
		OL CHIKI DIGIT TWO
		OL CHIKI DIGIT THREE
		OL CHIKI DIGIT FOUR
		OL CHIKI DIGIT FIVE
		OL CHIKI DIGIT SIX
		OL CHIKI DIGIT SEVEN
		OL CHIKI DIGIT EIGHT
		OL CHIKI DIGIT NINE
		OL CHIKI MU TTUDDAK
		OL CHIKI KAHLA TTUDDAK
		OL CHIKI MU-KAHLA TTUDDAK
		OL CHIKI RELA
		OL CHIKI PHARKA
		OL CHIKI OHOTT

Plane 00

Row xx

Introduction

The Ol Chiki script was invented by Pandit Raghunath Murmu (1905 - 1982) in the first half of the 20th century. The meaning of “Ol Chiki” is “the writing symbol” or “the writing script”. This script is also known as “*Ol Cemet*”, “*Ol Script*”, “*Ol Ciki Script*”, or simply as “*Ol*”. In Santali, “*Ol*” means “writing” and “*Cemet*” means “learning”. So, “*Ol Cemet*” means “The learning of writing”. Indeed, “*Ol Cemet*” (Murmu, 1976) is the title of the book, written by Pandit Raghunath Murmu for teaching the Ol Chiki script to the beginners. Santali is a language with its own special characteristics, and has a literature which dates back to the beginning of the 15th century (Murmu, 2002) and since then, it has been in various phases of its development in the subsequent centuries. Language is an instrument of group culture and identity. The script is a mechanism to propagate and preserve a language, which is necessary for the very survival of cultural communities. Naturally, he felt that Santals with their rich cultural heritage and tradition, also need a separate script to preserve and promote their language. Towards this goal, he wrote over 150 books covering wide spectrum of subjects such as grammar, novel, drama, poetry, and story in Ol Chiki as a part of his extensive programme for culturally upgrading the Santal community. “*Darege Dhan*”, “*Sidu-Kanhu*”, “*Bidu Chandan*” and “*Kherwal Bir*” are among the most acclaimed of his works. Later, he was honored by the Orissa and West Bengal Government.

Origin

Ol Chiki is alphabetic, and does not share any of the syllabic properties of other Indic scripts. It is used for writing Santali, which belongs to the Munda group of languages of the Austro-Asiatic family (Majumder, 2001). One of the interesting feature of the Ol Chiki script is that, it makes use of signs and symbols long familiar to the Santals. The very ingenuity in shaping the symbols of the letters and arranging the letters in the script has been greatly helpful in transmission of the script. A large number of words in the Santali language are derived from natural sounds. Letters of Ol Chiki script are also derived from the physical environment and what surrounds the people - hills, rivers, trees, birds, bees, plough, sickle - the list is endless (Mahapatra & Mohapatra, 2001). Zide in his work (Zide, 1996) says, “The shapes of the letters are not arbitrary, but reflect the names for the letters, which are words, usually the names of objects or actions represented—in conventionalized form – in the pictorial shape of the characters”. For example, the word 𑒧𑒱/ot/ means earth and the shape of letter O/AT/ is derived from the round shape of earth. Similarly 𑒧𑒱/ut/ means mushroom and so looks the shape of the letter 𑒧𑒱/UD/.

Explanations for the pictorial origin (Figure 1) of the letters in the 𑒧/E/ row are as follows: 𑒧/E/ “swelling”; 𑒧/EP/ “to meet, collide”; 𑒧/EDD/ “to point out a place”, a symbols (previously) used to “meeting place”; 𑒧/EN/ “to thrash grain” derived from a picture of two leg thrashing; 𑒧/ERR/ “to avoid” derived from a picture of a path that turns to avoid an obstruction or a danger (Zide, 1996).



Figure 1. Pictorial origin of the letters in the Z (e) row of Ol Chiki (Zide, 1996)

Major Users

“Santali” which literally means ‘the language of the Santals’ is spoken by the Santals who are also one of the oldest ethnic group in South Asia. Anthropologists tend to identify the Santals in the racial category of proto-australoid (Majumder, 2001). They are numerically the largest indigenous group in India having homogeneous characteristics and a total population of over 10 million (according to recent figures). The Santals are mostly concentrated in the Indian state of Jharkhand, Orissa, West Bengal and Assam, but they are also spread over sparsely in the neighboring country of Bangladesh and Nepal. Presently, the Santals have a high degree of bilingualism as they also speak the Indo-Aryan languages of the dominant population (viz..., Assamese, Bengali, Hindi, and Oriya as the case may be) besides Santali.

Santali Literatures

In earlier times, all Santali writings were in Bengali, Devnagari, Oriya or Roman script. Although there have been impressive works by foreigners and non-Santal writers on dictionary, collection of folklore etc., their work is mostly confined to research. Meanwhile, Roman script was in extensive use for writing Santali and several books in Santali have been published using Roman script. But most of the creative writings were written by the native speakers in Bengali, Devnagari or Oriya scripts. The use of different scripts for Santali has hindered the development and utilisation of Santali language. This, in turn, has effectively marred the progress of Santali language in several fields such as philosophy, history, religion, drama, novel, prose and poetry. The problem of using different scripts for the same language had necessitated the invention of a new script for Santali, and it finally led to the invention of Ol Chiki by Pandit Raghunath Murmu.

After the invention of Ol Chiki, a large number of books have been written by various authors in Santali using Ol Chiki script. Types of books include (i) novels and short stories, (ii) poetries, songs, and religious sermons, (iii) books on Santal society, (iv) primary books for learning Ol Chiki, (v) books for learning primary mathematics, (vi) books on Santali grammars and related topics, and (vii) books on great tribal persons. Santali magazines in Ol Chiki are also being published regularly. A partial list of such books and magazines is given in appendix E as a bibliography. Samples of front pages of magazines are given in appendix F.

Ol Chiki in Education

Private initiative of average Santals, co-operation and active support of various units of ASECA (Adivasi Socio-Educational and Cultural Association), a Santal organisation, made amazing inroads into rural Santal community for Ol Chiki and made considerable place for it among them. The footprints of proliferation of Ol Chiki among the Santals can be observed even in the remote villages of Nepal, far away from Santal heartland (The Kathmandu Post, dated January 15, 2001, Kathmandu). New era for Ol Chiki started after getting early recognition from the Govt. of West Bengal way back in 1979 (Appendix A). Recently, the Govt. of West Bengal in its extraordinary resolution published in the Calcutta Gazette dated 5th March 2001 (Appendix B) has constituted a committee “to study the feasibility and possibility of teaching Santali language using Ol Chiki script and to introduce it in various syllabus/curriculum starting from the primary school level up to the university level”. Now, it is the Jharkhand Government which gave green signal for teaching Santali language in Ol Chiki (Appendix C) as a separate subject starting from the primary school level up to secondary level. Apart from these developments for Ol Chiki, the Government of Orissa also recognises it for teaching Santali language in Ol Chiki in certain selected schools at the primary level, and for providing financial help to the ASECAs for teaching Santali language in Ol Chiki to all Santals, young and old alike.

Higher Education

The status of Ol Chiki in higher education is also equally impressive. Bihar Secondary School Examination Board, Patna and Jharkhand Secondary School Examination Board, Ranchi have accepted Santali Language/Literature as optional paper at the Matriculation level. Jharkhand Intermediate Education Board recently introduced compulsory *Matri Bhasa* (native language) subject where Santali carries equal weightage with Hindi. The Bihar Public Service Commission also offers Santali Language/Literature as optional paper for its competitive Examination. In West Bengal, the renowned Vishwa Bharati University, i.e., Shantiniketan has been imparting education in Santali language for the last 24 years or so. The Universities of Jharkhand and Bihar, viz., Ranchi University, Ranchi, Vinoba Bhave University, Hazaribagh, Sidu-Kanhu University, Dumka, and Baba Tilka Majhi University, Bhagalpur offer courses in Santali language/literature at the Post-Graduate level, and also conduct examinations for the same (Murmu, 2002). Appendix D gives a partial list of people who work in Santali language at the University/College level and also belong to the Santal community.

Ol Chiki Letters

The Ol Chiki letters are arranged in 6 by 5 matrix, in which the six letters in the first column of the matrix are vowels, and the remaining 24 letters are consonants. There are five basic diacritics in Ol Chiki, viz., *Mu Tudag* (◌̣), *Gahla Tudag* (◌̤), *Rela* (◌̥), *Pharka* (◌̦) and *Ahad* (◌̧). In Ol Chiki, diacritics have nice and interesting properties, and this is specially true with *Ahad*. The combination of *Mu Tudag* and *Gahla Tudag* results in *Mu-Gahla Tudag* (◌̣̤) which is also treated as a separate diacritic, and therefore, it is represented separately.

Majumder, Partha, 2001, "Ethnic populations of India as seen from an evolutionary perspective", *Journal of Bio-Science*, Vol-26, No.-4(Suppl), Indian Academy of Science.

Mohapatra, Shyam Sundar, 2001, "Formation of Ol Chiki script and process of its transmission", in *Santhal Worldview* edited by Nita Mathur, Indira Gandhi National Centre for the Arts, New Delhi.

Minegishi, Makoto & Murmu, Ganesh ,2001, *Santali Basic Lexicon with Grammatical Notes*. Tokyo: Institute for the Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies (Asian and African Lexicon Series No. 38).

Murmu, Ganesh, 2002, "Development of Santali Language and Literature and its Recognition(Language/Script movement)", In Proc. of the First National Seminar on Linguistic Landscaping in India with particular reference to the new States (21-22, February 2002), Central Institute of Indian Languages, Mysore - 570 006, India.

Murmu, Raghunath, 1976, ଓଡ଼ ଉଠୁଠୁଠ(Ol Cemet' – an Ol Chiki learning book for beginner), ASECA, Rairangpur, Orissa, India.

Murmu, Raghunath, 1976, ଚିକି ଗ୍ରନ୍ଥ(Ranal – A Santali Grammar in Santali), ASECA, Rairangpur, Orissa, India.

Murmu, Raghunath, 1966, ଦାରେଢ଼ନ ଦାୟତା(*Darege Dhan, Health is Wealth*), Published by Rukmani Murmu etc., Dandbose, Rairangpur, Orissa, India.

Soren, Nimai Charan, 2000, ଚିକି ମାଚେଟ(Chiki Mchet), Nohaini Memorial Education Award Committee, Rairangpur Orissa, India.

Zide, Norman, 1996 , "Script for Munda languages", in *The world's writing systems* edited by Peter T. Daniels and William Bright, Oxford University Press, New York.