

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

Doc Type: Working Group Document

Title: Proposal for encoding the Sorang Sompeng script in the BMP of the UCS

Source: Michael Everson, EGT (IE)

Status: Expert Contribution

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

Date: 1999-01-29

A. Administrative

1. Title

Proposal for encoding the Sorang Sompeng script in the BMP of the UCS.

2. Requester's name

Michael Everson, EGT (WG2 member for Ireland).

3. Requester type

Expert contribution.

4. Submission date

1999-01-29.

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. Sorang Sompeng.

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

No.

2. Number of characters

34

3. Proposed category

Category A.

4. Proposed level of implementation and rationale

Level 1 noncombining.

5a. Character names included in proposal?

Yes.

5b. Character names in accordance with guidelines?

Yes.

5c. Character shapes reviewable?

Yes (see below).

6a. Who will provide computerized font?

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?

No.

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. We need to contact Norman Zide and get his opinion of this proposal.

2. Information on the user community?

Speakers of the Sora language (population 273,911, according to the SIL *Ethnologue*).

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

To write the Sora language. Latin, Telugu, and Oriya scripts have also been used to write Sora.

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.

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?

No.

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

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

No.

E. Proposal

The Sorang Sompeng script is used to write the Sora language. Sora is a member of the Munda family of languages, which, together with the Mon-Khmer languages, makes up Austro-Asiatic. The Sora people live between the Oriya- and Telugu-speaking populations in what is now the Orissa-Andhra border area. Sorang Sompeng was devised by Mangei Gomango, son of charismatic community leader Malia Gomango, as part of a comprehensive cultural program, and was offered as an improvement over scripts used by Christian missionary linguists. Sorang Sompeng is used in primary and adult education, and is published in a variety of printed materials.

Structure

Sora appears to be alphabetic. Consonant clusters are written not with conjunct characters (as in Brahmic scripts) but by simply juxtaposing consonant letters, so the reader must recognize the presence or absence of the inherent schwa in any particular instance. There is no visible *virama* diacritic.

Names and ordering

The letters get their names from 24 gods in the Sora pantheon, e.g. *s* for Sundañ, *t* for Tənod; no rationale is apparent for the ordering. Consonant letter names have are derived by adding *a*’ to the consonant.

Unicode Character Properties

Spacing letters, category “Lo”, bidi category “L” (strong left to right)

xx00 - xx17

Numbers, decimal digits, category “Nd”, bidi category “L” (strong left to right)

xx20-xx29

Bibliography

Zide, Norman. 1996. “Scripts for Munda languages”, in Peter T. Daniels and William Bright, eds. *The world’s writing systems*. New York; Oxford: Oxford University Press. ISBN 0-19-507993-0

Zide gives other sources in his bibliography, none of which I have seen.

TABLE XXX - Row xx: SORANG SOMPENG

	xx0	xx1	xx2
0	ꞑ	Ꞓ	ꞓ
1	ꞔ	ꞕ	Ꞗ
2	ꞗ	Ꞙ	ꞙ
3	Ꞛ	ꞛ	Ꞝ
4	ꞝ	Ꞟ	ꞟ
5	Ꞡ	ꞡ	Ꞣ
6	ꞣ	Ꞥ	ꞥ
7	Ꞧ	ꞧ	Ꞩ
8	ꞩ		Ɡ
9	Ɬ		Ɪ
A	ꞯ		
B	Ʞ		
C	Ʇ		
D	Ʝ		
E	Ꭓ		
F	Ꞵ		

G = 00
P = 00

TABLE XXX - Row xx: SORANG SOMPENG

dec	hex	Name
	00	SORANG SOMPENG LETTER SA
	01	SORANG SOMPENG LETTER TA
	02	SORANG SOMPENG LETTER BA
	03	SORANG SOMPENG LETTER CA
	04	SORANG SOMPENG LETTER DA
	05	SORANG SOMPENG LETTER GA
	06	SORANG SOMPENG LETTER MA
	07	SORANG SOMPENG LETTER NGA
	08	SORANG SOMPENG LETTER LA
	09	SORANG SOMPENG LETTER NA
	0A	SORANG SOMPENG LETTER VA
	0B	SORANG SOMPENG LETTER PA
	0C	SORANG SOMPENG LETTER YA
	0D	SORANG SOMPENG LETTER RA
	0E	SORANG SOMPENG LETTER HA
	0F	SORANG SOMPENG LETTER KA
	10	SORANG SOMPENG LETTER JA
	11	SORANG SOMPENG LETTER NYA
	12	SORANG SOMPENG LETTER A
	13	SORANG SOMPENG LETTER E
	14	SORANG SOMPENG LETTER I
	15	SORANG SOMPENG LETTER U
	16	SORANG SOMPENG LETTER O
	17	SORANG SOMPENG LETTER AE
	18	(This position shall not be used)
	19	(This position shall not be used)
	1A	(This position shall not be used)
	1B	(This position shall not be used)
	1C	(This position shall not be used)
	1D	(This position shall not be used)
	1E	(This position shall not be used)
	1F	(This position shall not be used)
	20	SORANG SOMPENG DIGIT ZERO
	21	SORANG SOMPENG DIGIT ONE
	22	SORANG SOMPENG DIGIT TWO
	23	SORANG SOMPENG DIGIT THREE
	24	SORANG SOMPENG DIGIT FOUR
	25	SORANG SOMPENG DIGIT FIVE
	26	SORANG SOMPENG DIGIT SIX
	27	SORANG SOMPENG DIGIT SEVEN
	28	SORANG SOMPENG DIGIT EIGHT
	29	SORANG SOMPENG DIGIT NINE
	2A	(This position shall not be used)
	2B	(This position shall not be used)
	2C	(This position shall not be used)
	2D	(This position shall not be used)
	2E	(This position shall not be used)
	2F	(This position shall not be used)

Group 00

dec	hex	Name
-----	-----	------

Plane 00

Row xx