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 Shavian script in the SMP of the UCS**
Source:	**Michael Everson and John Jenkins**
Status:	**Expert contribution**
Action:	**For adoption by JTC1/SC2/WG2**
Date:	**2001-07-14**

This is a revision of L2/97-103, and contains the proposal summary form.

Background. The playwright George Bernard Shaw (1856-1950) was an outspoken critic of the idiosyncrasies of English orthography. In his will, he directed that Britain's Public Trustee seek out and publish an alphabet of no fewer than forty letters to provide for the phonetic spelling of English. The alphabet finally selected was designed by Kingsley Read and is variously known as Shavian, Shaw's alphabet, and the Proposed British Alphabet.

Also in accordance with Shaw's will, an edition of his play, *Androcles and the Lion* was published and distributed to libraries, containing the text in both the standard Latin alphabet and in Shavian.

As with other attempts at spelling reform in English, the alphabet has met with little success. It has, nonetheless, its advocates and users. There are at least two implementations for computers, for example, and a standard implementation of the private use area includes an encoding of Shavian. The normative version of Shavian is taken to be the version in *Androcles and the Lion*.

Structure. The alphabet consists of forty-eight letters and one punctuation mark. The letters have no case. The digits and other punctuation marks are the same as Latin. The one additional punctuation mark is a "name mark," used to indicate proper nouns. The letter names are intended to be indicative of their sounds; thus the sound /p/ is represented by the SHAVIAN LETTER PEEP (𐑐).

The first forty letters are divided into four groups of ten. The first ten and second ten are 180-degree rotations of one another; letters of the third and fourth often show a similar relationship of shape.

The first ten letters are "tall" letters (which ascend above the x-height) and generally represent unvoiced consonants. The next ten letters are "deep" letters (which descend below the baseline) and generally represent voiced consonants. The next twenty are the vowels and liquids. Again, each of these letters usually has a close phonetic relationship to the letter in its matching set of ten.

The remaining eight letters are technically ligatures, the first six involving vowels plus /r/. Since ligation isn't optional, these eight letters are included in the proposal.

Proposal. This proposal is based on the version of the Shaw alphabet found in the private-use ConScript Unicode Registry. It allocates three columns for the forty-eight letters and uses U+00B7 MIDDLE DOT for the name mark. The Unicode Technical Committee has approved this proposal.

The proposed location for the encoding of the Shaw alphabet is the SMP, and in the same block as the Deseret Alphabet. Shavian's limited use makes this a reasonable location. This is also the position indicated for Shavian in the current Road Map proposal.

Issues. As with the Deseret Alphabet, the problem of collation is not addressed by the alphabet's designers.

References

Cowan, John, and Michael Everson. 1997. *Shavian ConScript Unicode Standard.* http://www.egt.ie/standards/csur/shavian.html.
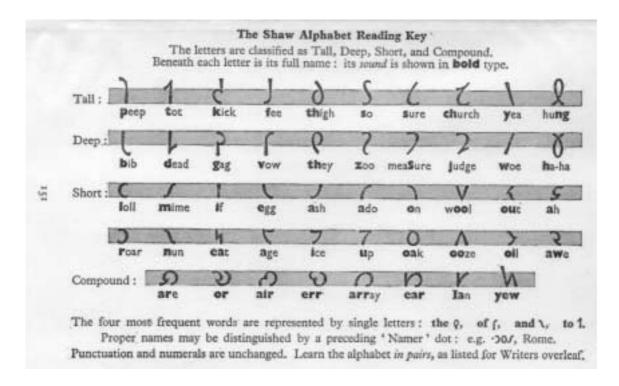
Crystal, David, ed. 1987. *The Cambridge Encyclopedia of Language.* Cambridge: University Press.

DeMeyere, Ross. 1997. *About Shavian.* http://www.demeyere.com/Shavian/info.html.

Shaw, George Bernard. 1962. *Androcles and the Lion; an Old Fable Renovated by Bernard Shaw, with a Parallel Text in Shaw's Alphabet to be Read in Conjunction Showing its Economies in Writing and Reading.* Harmondsworth: Penguin Books

Examples

A key to the alphabet from Androcles and the Lion:

The opening of *Androcles and the Lion*:
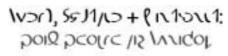

ANDROCLES AND THE LION

PROLOGUE

Overture: forest sounds, roaring of lions, Christian hymn faintly.

A jungle path. A lion's roar, a melancholy suffering roar, comes from the jungle. It is repeated nearer. The lion limps from the jungle on three legs, holding up his right forepaw, in which a huge thorn sticks. He sits down and contemplates it. He licks it. He shakes it. He tries to extract it by scraping it along the ground, and hurts himself worse. He roars piteously. He licks it again. Tears drop from his eyes. He limps painfully off the path and lies down under the trees, exhausted with pain. Heaving a long sigh, like wind in a trombone, he goes to sleep.
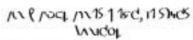
A key to the alphabet from Androcles and the Lion:

The announcement for the Tenth International Unicode Conference in Shavian:

Unicode Character Properties

Spacing letters, category "Lo" (other), bidi category "L" (strong left to right)

10450-1047F

A. Administrative

1. Title

Revised proposal for encoding the Shavian script in the SMP of the UCS

2. Requester's name

Michael Everson and John Jenkins

3. Requester type

Expert contribution

4. Submission date

2001-07-14

5. Requester's reference

6a. Completion

This is a complete proposal. It has been stable since 1997. It has been approved by the UTC.

6b. More information to be provided?

No

B. Technical -- General

1a. New script? Name?

Yes. Shavian.

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

No.

2. Number of characters

48

3. Proposed category

Category D

4. Proposed level of implementation and rationale

Level 1. No combining characters.

5a. Character names included in proposal?

Yes

5b. Character names in accordance with guidelines?

Yes

5c. Character shapes reviewable?

Yes

6a. Who will provide computerized font?

Michael Everson, Everson Typography

6b. Font currently available?

Yes

6c. Font format?

TrueType

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

Yes. See bibliography.

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

Yes. See below.

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

Yes.

C. Technical -- Justification

1. Contact with the user community?
Yes, such as it is.
2. Information on the user community?
Students of English orthography, Shaw, writing systems, and script enthusiasts.
3a. The context of use for the proposed characters?
Proposed for use to write the English language. Was only used to publish one book, but the script is cited in many works. A Google search on "Shavian alphabet", "Shaw's alphabet", and "Proposed British Alphabet" yields about 250 hits.
3b. Reference
4a. Proposed characters in current use?
Shavian enjoys some use and there are numerous computer implementations (fonts) available.
4b. Where?
Chiefly on the internet.
5a. Characters should be encoded entirely in BMP?
No.
5b. Rationale
In keeping 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?
No.
9b. List of composite sequences and their corresponding glyph images provided?
No.
10. Characters with any special properties such as control function, etc. included?
No

D. SC2/WG2 Administrative

To be completed by SC2/WG2
1. Relevant SC 2/WG 2 document numbers:

2. Status (list of meeting number and corresponding action or disposition)

3. Additional contact to user communities, liaison organizations etc.

4. Assigned category and assigned priority/time frame

Other Comments

Row 104: SHAVIAN

	1045	1046	1047
0			
1			
2			
3			
4			
5			
6			
7			
8			
9			
A			
B			
C			
D			
E			
F			

G = 00
P = 01

Row 104: SHAVIAN

hex	Name	hex	Name
50	SHAVIAN LETTER PEEP		
51	SHAVIAN LETTER TOT		
52	SHAVIAN LETTER KICK		
53	SHAVIAN LETTER FEE		
54	SHAVIAN LETTER THIGH		
55	SHAVIAN LETTER SO		
56	SHAVIAN LETTER SURE		
57	SHAVIAN LETTER CHURCH		
58	SHAVIAN LETTER YEA		
59	SHAVIAN LETTER HUNG		
5A	SHAVIAN LETTER BIB		
5B	SHAVIAN LETTER DEAD		
5C	SHAVIAN LETTER GAG		
5D	SHAVIAN LETTER VOW		
5E	SHAVIAN LETTER THEY		
5F	SHAVIAN LETTER ZOO		
60	SHAVIAN LETTER MEASURE		
61	SHAVIAN LETTER JUDGE		
62	SHAVIAN LETTER WOE		
63	SHAVIAN LETTER HA-HA		
64	SHAVIAN LETTER LOLL		
65	SHAVIAN LETTER MIME		
66	SHAVIAN LETTER IF		
67	SHAVIAN LETTER EGG		
68	SHAVIAN LETTER ASH		
69	SHAVIAN LETTER ADO		
6A	SHAVIAN LETTER ON		
6B	SHAVIAN LETTER WOOL		
6C	SHAVIAN LETTER OUT		
6D	SHAVIAN LETTER AH		
6E	SHAVIAN LETTER ROAR		
6F	SHAVIAN LETTER NUN		
70	SHAVIAN LETTER EAT		
71	SHAVIAN LETTER AGE		
72	SHAVIAN LETTER ICE		
73	SHAVIAN LETTER UP		
74	SHAVIAN LETTER OAK		
75	SHAVIAN LETTER OOZE		
76	SHAVIAN LETTER OIL		
77	SHAVIAN LETTER AWE		
78	SHAVIAN LETTER ARE		
79	SHAVIAN LETTER OR		
7A	SHAVIAN LETTER AIR		
7B	SHAVIAN LETTER ERR		
7C	SHAVIAN LETTER ARRAY		
7D	SHAVIAN LETTER EAR		
7E	SHAVIAN LETTER IAN		
7F	SHAVIAN LETTER YEW		