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: 2001-06-03

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 ( ).

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.

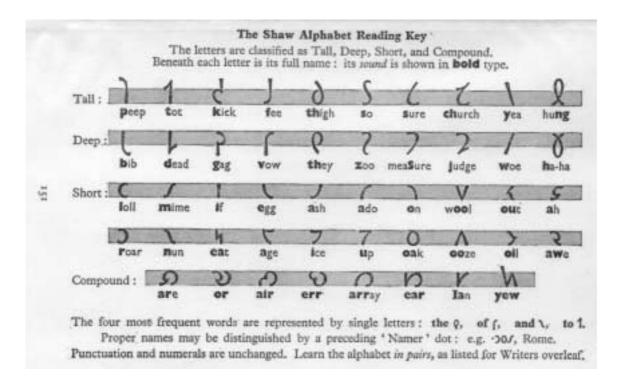
Crystal, David, ed. 1987. The Cambridge Encyclopedia of Language. Cambridge: University Press.

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*:

## JULOCCOS V 6 CALV

## J20C/5

[סלטנאט: אאנן זכאלי אונן לבעלי קונן-

( 318/40 ) 49. ( 224/5) 29. ( 1404/40 ) 31/414 ) 39. (41/4 ) 31/414 ) 39. (41/4 ) 31/4 ( 278/40 ) 41/4 ) 41/4 ( 31/4 ) 41/4 ) 41/4 ( 31/4 ) 41/4 ) 41/4 ( 31/4 ) 41/4 ) 41/4 ( 31/4 ) 41/4 ) 41/4 ( 31/4 ) 41/4 ) 41/4 ( 31/4 ) 41/4 ) 41/4 ( 31/4 ) 41/4 ) 41/4 ) 41/4 ( 31/4 ) 41

#### ANDROCLES AND THE LION

#### **PROLOGUE**

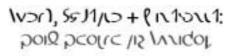
Overture: forest sounds, roaring of lions, Christian hymn faintly.

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.

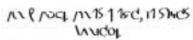
A key to the alphabet from Androcles and the Lion:

Doubl	a Umas	- 1	etween pair	w show the	rel	ati	ue he	ight of
Talls, letters	Deeps,	, ai	nd Shorts. is; those ee heading	Wherever starred *	r p	DSS []	ible, be v	finish written
PER ST	Tall D	eep	to seemed	Sh	ort		Short	
реер	):1	(	bib	<b>If</b>	1	:	4	eat
<b>t</b> ot	1:	1	dead	egg	1	:	۲	age
<b>k</b> ick	9:	9	gag	ash*	J		7	fce
fee	J:	1	<b>Vow</b>	ado*	1	:	7	up
thigh	9:	9	they	<b>o</b> n	1		0	oak
so	5:	2	200	wool	٧		٨	<b>00</b> Ze
Sure	4:	7	meaSure	out	4		>	oll
church	1:	2	<b>J</b> udge	ah*	5		5	awe
yea	1:	/	*Woe	are	n	:	D	or
hung	2:	ď	tha-ha	air	D		v	err
	Short S	Shor	t	array	0	:	n	ear
foll	C:	)	Foar				Tall	
-	1:	1	nun	Ian	r	-	W	yew

The announcement for the Tenth International Unicode Conference in Shavian:



agéto vi ka (tinà ntorvario Invido, dintons, fili gias, in viac 10-12, 1997, in vaio, garrini (dintons) ja gall laggio nysta-jas, udstats in (paogra ntoni) (tinidos, in torvario galani) codra galani, intervitados ( tinidos in staetla sestina (tinidados), tinis, sudst curl, (tinidados dir tinia).



Evilan againgt very cost (inspector/large), authoritings agenty this cost (installed profession).

## **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

2001-05-03

## 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	)	7	Ч	
1	1	2	7	
2	4	/	フ	
3	J	Q	7	
4	9	С	0	
5	S	5	٨	
6	2	1	>	
7	Z	Ĺ	5	
8	\	J	S	
9	Q	(	ઇ	
Α	l	)	$\mathcal{C}$	
В	L	V	S	
c 2		<	$\sim$	
D		9	2	
Е	6	C	r	
F	7	\	W	

G = 00 P = 01

## **Row 104: SHAVIAN**

hex	Name	hex	Name
50 51 52 53 55 55 56 57 58 59 58 50 50 61 62 63 64 66 66 66 67 77 77 78 77 78 77 77 77 77 77 77 77 77	SHAVIAN LETTER PEEP SHAVIAN LETTER TOT SHAVIAN LETTER KICK SHAVIAN LETTER FEE SHAVIAN LETTER SO SHAVIAN LETTER SURE SHAVIAN LETTER SURE SHAVIAN LETTER SURE SHAVIAN LETTER CHURCH SHAVIAN LETTER HUNG SHAVIAN LETTER PEAD SHAVIAN LETTER WOW SHAVIAN LETTER MEASURE SHAVIAN LETTER MEASURE SHAVIAN LETTER BEAD SHAVIAN LETTER BEAD SHAVIAN LETTER WOS SHAVIAN LETTER WOS SHAVIAN LETTER WOS SHAVIAN LETTER BEAD SHAVIAN LETTER BEAD SHAVIAN LETTER WOS SHAVIAN LETTER BOD SHAVIAN LETTER BOD SHAVIAN LETTER ADD SHAVIAN LETTER AND SHAVIAN LETTER AND SHAVIAN LETTER AND SHAVIAN LETTER WOOL SHAVIAN LETTER ON SHAVIAN LETTER ROR SHAVIAN LETTER BOD SHAVIAN LETTER ROR SHAVIAN LETTER BOD SHAVIAN LETTER BOS SHAV		