Accredited Standards Committee[1]

NCITS, Information Processing Systems

Doc:	**L2/97-103**
Date:	21 May 1997
Author:	John H. Jenkins
	Apple Computer, Inc.
	2 Infinite Loop
	M/S 302-2IS
	Cupertino CA 95014
Phone:	(408) 974-6276
Fax:	(408) 862-4566
e-mail:	jenkins@apple.com

Proposal to add Shavian to Unicode/ISO-IEC 10646

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 Unicode's user zone 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 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 ConScript Unicode user-zone registry. It allocates three columns for the forty-eight letters and uses U+00B7 MIDDLE DOT for the name mark.

The proposed location for the encoding of the Shaw alphabet is on Plane 1, off the BMP, 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. We indicate in **boldface** sound in each letter's name which the letter represents. This is a convention adopted from *Androcles and the Lion* but which need not be maintained in Unicode and ISO/IEC 10646.

0001 xx00	)	SHAVIAN LETTER **P**EEP
0001 xx01	1	SHAVIAN LETTER **T**OT
0001 xx02	⌐	SHAVIAN LETTER **K**ICK
0001 xx03	⌡	SHAVIAN LETTER **F**EE
0001 xx04	∂	SHAVIAN LETTER **TH**IGH
0001 xx05	∫	SHAVIAN LETTER **S**O
0001 xx06	⌐	SHAVIAN LETTER **S**URE
0001 xx07	⌐	SHAVIAN LETTER **CH**URCH
0001 xx08	\	SHAVIAN LETTER **Y**EA
0001 xx09	ℓ	SHAVIAN LETTER HU**NG**
0001 xx0A	(	SHAVIAN LETTER **B**IB
0001 xx0B	)	SHAVIAN LETTER **D**EAD
0001 xx0C	⍴	SHAVIAN LETTER **G**AG
0001 xx0D	ſ	SHAVIAN LETTER **V**OW
0001 xx0E	℘	SHAVIAN LETTER **TH**EY
0001 xx0F	⌐	SHAVIAN LETTER **Z**OO
0001 xx10	⌐	SHAVIAN LETTER MEA**S**URE
0001 xx11	2	SHAVIAN LETTER **J**UDGE
0001 xx12	/	SHAVIAN LETTER **W**OE
0001 xx13	∂	SHAVIAN LETTER **H**A-HA
0001 xx14	c	SHAVIAN LETTER **L**OLL
0001 xx15	⌐	SHAVIAN LETTER **M**IME
0001 xx16	ı	SHAVIAN LETTER **I**F

0001 xx17	ᴕ	SHAVIAN LETTER **E**GG
0001 xx18	ᴄ	SHAVIAN LETTER **A**SH
0001 xx19	ᴊ	SHAVIAN LETTER **A**DO
0001 xx1A	ᴝ	SHAVIAN LETTER **O**N
0001 xx1B	ᴠ	SHAVIAN LETTER W**OO**L
0001 xx1C	ᴋ	SHAVIAN LETTER **OU**T
0001 xx1D	ᴣ	SHAVIAN LETTER **A**H
0001 xx1E	ᴐ	SHAVIAN LETTER **R**OAR
0001 xx1F	ᴌ	SHAVIAN LETTER RU**N**
0001 xx20	ᴕ	SHAVIAN LETTER **EA**T
0001 xx21	ᴛ	SHAVIAN LETTER **A**GE
0001 xx22	ᴚ	SHAVIAN LETTER **I**CE
0001 xx23	ᴦ	SHAVIAN LETTER **U**P
0001 xx24	ᴏ	SHAVIAN LETTER **OA**K
0001 xx25	ᴧ	SHAVIAN LETTER **OO**ZE
0001 xx26	ᴕ	SHAVIAN LETTER **OI**L
0001 xx27	ᴚ	SHAVIAN LETTER **AWE**
0001 xx28	ᴕ	SHAVIAN LETTER **AR**E
0001 xx29	ᴕ	SHAVIAN LETTER **OR**
0001 xx2A	ᴕ	SHAVIAN LETTER **AIR**
0001 xx2B	ᴕ	SHAVIAN LETTER **ERR**
0001 xx2C	ᴕ	SHAVIAN LETTER **ARR**AY
0001 xx2D	ᴕ	SHAVIAN LETTER **EAR**
0001 xx2E	ᴧ	SHAVIAN LETTER **IA**N
0001 xx2F	ᴕ	SHAVIAN LETTER **YEW**

A table of the forty-eight proposed letters follows:

	0001 xx0	**0001 xx1**	**0001 xx2**
0	ᴉ	ᴚ	ᴕ
1	ᴉ	ᴚ	ᴛ
2	ᴉ	ᴵ	ᴚ
3	ᴊ	ᴕ	ᴦ
4	ᴕ	ᴄ	ᴏ
5	ᴕ	ᴕ	ᴧ

7	ʋ	⟝	⟨
8	⟍	⟨	⟁
9	⟒	⟌	⟂
A	⟊	⟍	⟃
B	⟎	∨	⟅
C	⟓	⟨	⌒
D	⟐	⟲	⟆
E	⟑	⟂	⟁
F	⟔	⟍	⟎

Issues

The problem of collation is not addressed by the alphabet's designers.

References

Cowan, John, and Everson, Michael. 1997. Shavian ConScript Unicode Standard. http://www.indigo.ie/egt/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 Shaw Alphabet Reading Key

The letters are classified as Tall, Deep, Short, and Compound.
Beneath each letter is its full name : its *sound* is shown in **bold** type.

Tall :									
peep	tot	kick	fee	thigh	so	sure	church	yea	hung

Deep :									
bib	dead	gag	vow	they	zoo	meaSure	judge	woe	ha-ha

| Short :									
loll	mime	if	egg	ash	ado	on	wool	out	ah
roar	nun	eat	age	ice	up	oak	ooze	oil	awe

Compound :							
are	or	air	err	array	ear	Ian	yew

The four most frequent words are represented by single letters : the ϙ, of ſ, and ↖, to ʇ.
Proper names may be distinguished by a preceding 'Namer' dot : e.g. ·ꓛ0ſ, Rome.
Punctuation and numerals are unchanged. Learn the alphabet *in pairs*, as listed for Writers overleaf.

The opening of *Androcles and the Lion*:

The announcement for the Tenth International Unicode Conference in Shavian (note the lack of the naming mark in the text—mea culpa, mea culpa, mea maxima culpa):

[Shavian script heading]

[Shavian script heading, in red]

[Shavian script paragraph, including:] 10-12, 1997, _[Shavian script continues]_

[Shavian script heading]

[Shavian script paragraph]

[Shavian script paragraph]

[Shavian script paragraph]

[Shavian script paragraph]

[Shavian script paragraph, including:] 25 _[Shavian script]_