

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

**Doc Type: Working Group Document****Title: Preliminary proposal for encoding the Aiha script in the SMP of the UCS****Source: Michael Everson****Status: Individual Contribution****Action: For consideration by JTC1/SC2/WG2 and UTC****Date: 2023-06-16**

**1. Introduction.** The Aiha script is used to write the Kesh language, devised by Ursula K. Le Guin for her novel *Always Coming Home* (1985). Le Guin is considered by many to be one of America's premier writers, philosophers, and agitators. Interest in Le Guinian studies appears to be growing, in a way analogous to the way J. R. R. Tolkien inspired Tolkienien studies.

My own interest in Kesh and Aiha is linguistic, and I am working on a book about Le Guin's efforts as a word-designer and a language-creator. This is, I think, valuable to those who are interested in other aspects of Ursula's poetic work and style, as well as to those who study or devise constructed languages and scripts. My research on the writing system prompts me to wish that the encoding of Aiha were formalized. This proposal reflects Kesh as described in the charts in the novel. This differs in some particulars from what is in the corpus, which shows some script development. A final proposal will be produced at the conclusion of that research.

**2. Description.** Aiha has nineteen consonant letters. Two of these behave as a kind of "casing" pair, a marked and unmarked form, equivalent but distinguished in writing and alphabet charts (✓ *gf ġ*, ~ *hw*, ~ *hw̄*). Aiha has eight vowels, four of which are given in alphabet charts where size is described as an explicit option and sometimes preference for users. Two of the Kesh spoken diphthongs are represented in writing by ligated but not decomposable letters.

**3. Processing.** Conventional LTR directionality is expected for Aiha.

"Text was written and read left to right, top to bottom; but clowns practiced reverse writing, bottom to top, right to left, the letters reversed."

Support for clowns would involve the same practices that would apply to Latin or Old Italic.

Consonant gemination is indicated by one of two diacritical marks; another diacritical mark might be placed on a vowel to indicate prolongation of the entire word. It may be the case that this is written long above the whole word itself. Further study is needed.

**4. Numerals.** No digits have been described. The Kesh names of numbers are known, but no notation is described or in evidence.

**5. Punctuation.** Aiha script has its own punctuation marks. The ones used syntactically are similar to ours. A single line — is used for a FULL STOP, and an angled one or “double period” ↵ is used to indicate a paragraph break. The COMMA is represented by a squiggle Ƶ and a tailed squiggle ↷ for a SEMICOLON.

**6. Character names.** The names used for the characters here are based on the chart; there are no other names known.

**7. Sorting.** The two “marked” consonant letters and four “small” vowel letters do not appear in the code chart pending further study. If encoded, they would sort as equivalents as shown below.

ƶ K < Ƨ G << Ƨ Ğ < Ƨ Š < Ƨ Ć < Ƨ L < Ƨ N < Ƨ S <  
 Δ D < Ƨ T < Ƨ R < Ƨ F < Ƨ V < Ƨ M < Ƨ B < · P <  
 Ƨ W < ~ HW << ~ HŴ < Ƨ Y < Ƨ H <  
 Ƨ O << Ƨ Q < Ƨ Ó << Ƨ Ǫ < Ƨ OU < Ƨ Ú << Ƨ Ū <  
 Ƨ E < Ƨ A << Ƨ Ā < Ƨ AI < Ƨ I < Ƨ Í < Ƨ Z < Ƨ

**8. Unicode Character Properties**

To be supplied

**9. Bibliography**

Le Guin, Ursula K. 1985. *Always Coming Home*. Composer: Todd Barton. Artist: Margaret Chodos. New York: Harper & Row. ISBN 0-06-015545-0.  
 ——. 2019. *Always Coming Home: Author’s Expanded Edition*. Composer: Todd Barton. Artist: Margaret Chodos. Edited by Brian Attebery. New York: Library of America. ISBN 978-1-59853-603-4.

**10. Figures.**

**PUNCTUATION.**

In inscriptions and mural writing little punctuation was used except for a slanting stroke to divide sentences. In ordinary and literary writing, punctuation was careful and complex, including indications of expression and tempo which we use only for music. The principal signs were:

- Equivalent to our period
- ↵ A “double period,” roughly equivalent to a paragraph break
- Ƶ Equivalent to our comma, indicating a phrase within a continuing sentence
- ↷ Equivalent to our semicolon, indicating a self-contained phrase within a continuing sentence

These four signs, like our punctuation, were syntactically meaningful and aided clarity. The next five concern dynamics and tempo:

- / Equivalent to our dash, signifying a pause. Repeated, a long pause; repeated more than once, a longer pause
- word Kesh underlining, just like ours, denotes emphasis or stress
- word̄ The opposite of underlining: de-emphasis, a soft or even tone
- ~ Written over a word, a fermata: prolong the word. Written in the margin: rallentando: read this line or these lines slowly
- ↵ Written in the margin: speed up, or resume normal reading pace

**Figure 1.** Discussion of punctuation marks from Le Guin 1985:498.

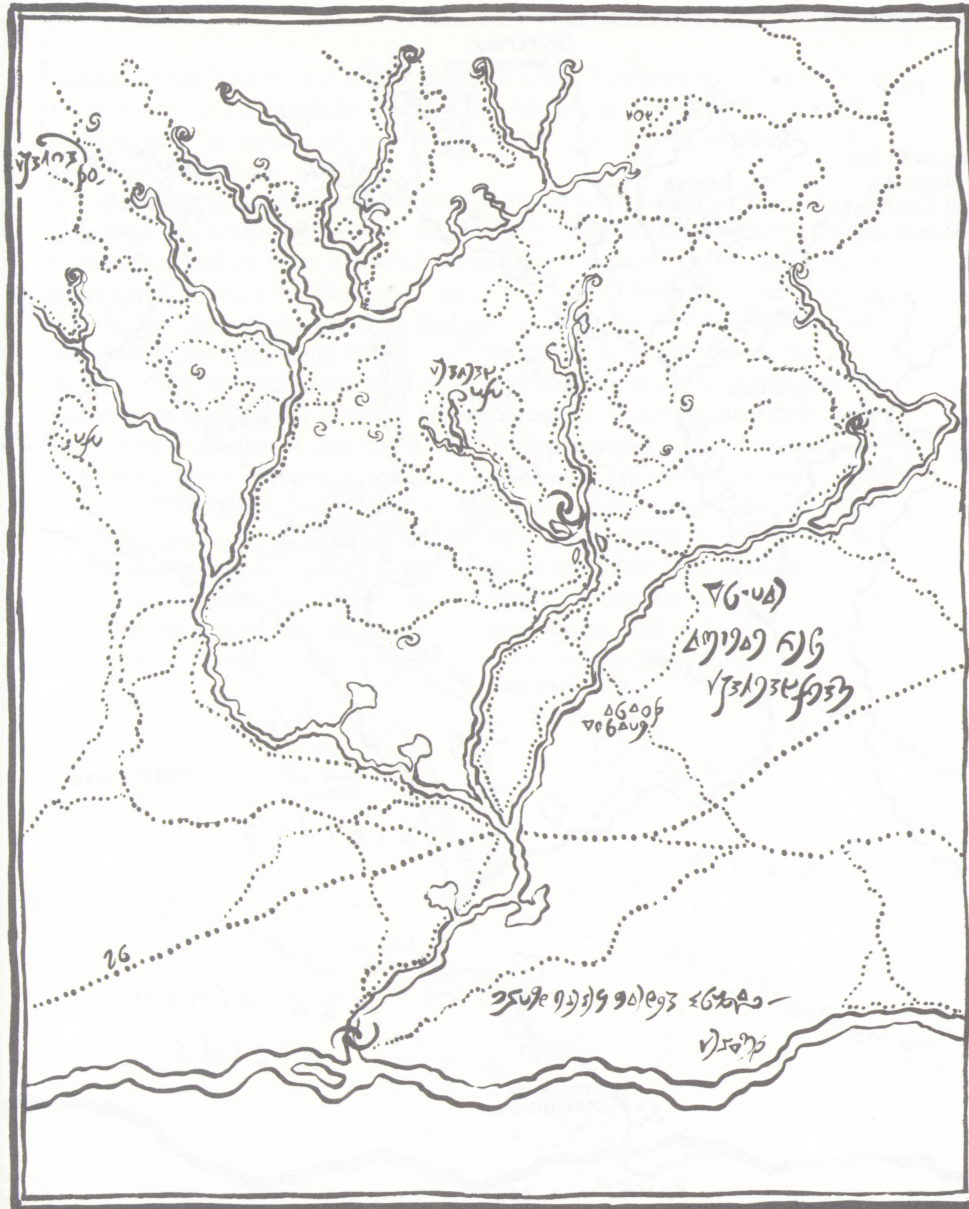
## The Kesh Alphabet

KESH ALPHABET	ENGLISH ALPHABET	INTERNATIONAL PHONETIC ALPHABET
	k	[k]
	g	[g]
	sh	[ʃ]
	ch	[tʃ]
	l	[l], [ɫ] (The two kinds of l in 'little')
	n	[n]
	s	[s]
	d	[d], [ɖ], [ð]
	t	[t]
	r	[r̄], [r̥], [dr], [ð] (See note below)
	f	[f]
	v	[v]
	m	[m]
	b	[b]
	p	[p]
	w	[w], [ʷ]
	hw	[hw] (As in English "what")
	y	[y], [ʏ]
	h	[h], sometimes [x]
	o	[ɔ] (As in English "off")
	ó	[o] (As in "oat" without glide)
	ou	[ow] (As in "go")
	ú	[u] (As in "toot")
	u	[ə]; [ʌ] (As in "the"; as in "but," "dumb")
	e	[ɛ] (As in "yet")
	a	[a] (As in "father")
	ai	[aʲ] (as in "tie")
	i	[ɪ] (As in "pit")
	í	[i] (As in "meet")
	Five-House sign, pronounced [z] (a suffix to words in the Five House mode, which is not used in most kinds of writing)	
	Four-House sign (there is no spoken sign)	
	Doubled letter sign, written over the letter.	

Note concerning the Kesh *r*: depending on context it may be a trill, a flap (as in English "steady" or "Betty"), the fricative [ð] as in "then," or a stop [dr]; and as a final sound it is often very like American English "hard *r*" in "her."










































**Figure 2.** Kesh alphabet chart from Le Guin 1985:496–497.
























**Figure 3.** Map from Le Guin 1985:451. The text of the map is given in the centre right and reads: *tupúde dajhada bóye sinšanv ġanaj* ‘Some of the Paths around Sinshan Creek’.













	F8A	F8B	F8C	F8D
0	 F8A0	 F8B0	 F8C0	
1	 F8A1	 F8B1	 F8C1	
2	 F8A2	 F8B2	 F8C2	
3	 F8A3	 F8B3	 F8C3	
4	 F8A4	 F8B4	 F8C4	
5	 F8A5	 F8B5	 F8C5	
6	 F8A6	 F8B6	 F8C6	
7	 F8A7	 F8B7	 F8C7	
8	 F8A8	 F8B8		
9	 F8A9	 F8B9		
A	 F8AA	 F8BA		
B	 F8AB	 F8BB		
C	 F8AC	 F8BC		
D	 F8AD	 F8BD		
E	 F8AE	 F8BE		
F	 F8AF	 F8BF	 F8CF	

The Aiha script was devised by Ursula K. Le Guin to write the Kesh language in her 1985 novel *Always Coming Home*.



### Consonants

F8A0		AIHA LETTER K
F8A1		AIHA LETTER G
F8A2		AIHA LETTER SH
F8A3		AIHA LETTER CH
F8A4		AIHA LETTER L
F8A5		AIHA LETTER N
F8A6		AIHA LETTER S
F8A7		AIHA LETTER D
F8A8		AIHA LETTER T
F8A9		AIHA LETTER R
F8AA		AIHA LETTER F
F8AB		AIHA LETTER V
F8AC		AIHA LETTER M
F8AD		AIHA LETTER B
F8AE		AIHA LETTER P
F8AF		AIHA LETTER W
F8B0		AIHA LETTER HW
F8B1		AIHA LETTER Y
F8B2		AIHA LETTER H

### Vowels

F8B3		AIHA LETTER O
F8B4		AIHA LETTER OO
F8B5		AIHA LETTER OU
F8B6		AIHA LETTER UU
F8B7		AIHA LETTER U
F8B8		AIHA LETTER E
F8B9		AIHA LETTER A
F8BA		AIHA LETTER AI
F8BB		AIHA LETTER I
F8BC		AIHA LETTER II





### House signs

F8BD		AIHA FIVE-HOUSE SIGN
F8BE		AIHA FOUR-HOUSE SIGN





### Symbol

F8BF		AIHA HEYTYA-IF
------	---	----------------

### Punctuation

F8C0		AIHA FULL STOP
F8C1		AIHA DOUBLE FULL STOP
F8C2		AIHA COMMA
F8C3		AIHA SEMICOLON

### Prosodic marks

F8C4		AIHA DASH
F8C5		AIHA DE-EMPHASIS SIGN
F8C6		AIHA PROLONGATION SIGN
F8C7		AIHA HASTENING SIGN

### Combining mark

F8CF		AIHA GEMINATION MARK
------	---	----------------------

## A. Administrative

1. Title

**Preliminary proposal for encoding the Aiha script in the SMP of the UCS**

2. Requester's name

**Michael Everson**

3. Requester type (Member body/Liaison/Individual contribution)

**Individual contribution.**

4. Submission date

**2023-06-16**

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

**No.**

6b. More information will be provided later

**Yes.**

## B. Technical – General

1. Choose one of the following:

1a. This proposal is for a new script (set of characters)

**Yes.**

1b. Proposed name of script

**Aiha.**

1c. The proposal is for addition of character(s) to an existing block

**No.**

1d. Name of the existing block

2. Number of characters in proposal

**41.**

3. Proposed category (A-Contemporary; B.1-Specialized (small collection); B.2-Specialized (large collection); C-Major extinct; D-Attested extinct; E-Minor extinct; F-Archaic Hieroglyphic or Ideographic; G-Obscure or questionable usage symbols)

**Category B-1.**

4a. Is a repertoire including character names provided?

**Yes.**

4b. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document?

**Yes.**

4c. Are the character shapes attached in a legible form suitable for review?

**Yes.**

5a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?

**Michael Everson.**

5b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.).

**Michael Everson.**

6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

**Yes.**

6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

**Yes.**

7. Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?

**Yes.**

8. Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see UAX #44 <http://www.unicode.org/reports/tr44/> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

**See above.**

## C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

**No.**

2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?

**Yes.**

2b. If YES, with whom?

**Theo Downes-Le Guin, representing the Ursula K. Le Guin Estate; Emily Floyd, who has worked artistically with Aiha.**

2c. If YES, available relevant documents

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:

**The people in the novel use the script natively. The user community for encoded characters would be small in comparison to some other scripts (whether used in fiction or otherwise).**

4a. The context of use for the proposed characters (type of use; common or rare)

**Rare enough.**

4b. Reference

5a. Are the proposed characters in current use by the user community?

**Yes.**

5b. If YES, where? Reference:

**To be supplied.**

6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?

**No.**

6b. If YES, is a rationale provided?

6c. If YES, reference

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

**Yes.**

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

**No.**

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

**No.**

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

**No.**

10b. If YES, is a rationale for its inclusion provided?

10c. If YES, reference

11a. Does the proposal include use of combining characters and/or use of composite sequences?

**Yes.**

11b. If YES, is a rationale for such use provided?

**Yes.**

11c. If YES, reference

**There is a combining character proposed.**

11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

**No.**

11e. If YES, reference

12a. Does the proposal contain characters with any special properties such as control function or similar semantics?

**No.**

12b. If YES, describe in detail (include attachment if necessary)

13a. Does the proposal contain any Ideographic compatibility character(s)?

**No.**

13b. If YES, is the equivalent corresponding unified ideographic character(s) identified? If YES, reference