

Proposal to encode the Kaithi Letter ZHA in Unicode

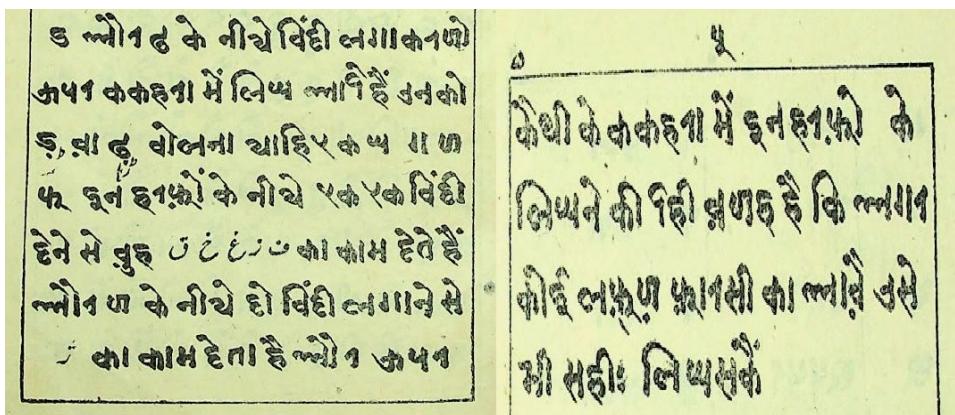
Biswajit Mandal (biswajitmandal.bm90@gmail.com)

Dec 16, 2025

This is a proposal to encode a new character in the Kaithi block of the Unicode standard:

GLYPH	CODE	CHARACTER NAME
କାଇଥି ଝା	U+110C3	KAITHI LETTER ZHA

Kaithi (କାଇଥି), also called Kayathi (କ୍ୟାଇଥି) is a Brahmic script historically used across parts of Northern and Eastern India. It was prevalent in regions corresponding to modern-day Uttar Pradesh, Bihar, and Jharkhand. The script was primarily utilized for legal, administrative, and private records and was adapted for a variety of Indo-Aryan languages, including Angika, Awadhi, Bhojpuri, Maithili, Magahi and etc. In Kaithi script, to write the Perso-Arabic loanwords they used ଙ Nuqta sign beneath the letters like ଖୁଁ=ଖୁ q, ଖୁଁ=ଖୁ x, ଖୁଁ=ଖୁ ଗୁଁ=ଗୁଁ ga, ଖୁଁ=ଖୁଁ ଜୁଁ=ଜୁଁ za, ଖୁଁ=ଖୁଁ ଫୁଁ=ଫୁଁ fa, etc. The proposed letter is also used to write a particular phoneme of the loanword i.e ଖୁଁ=ଜୁଁ za.



Translation: ୬ and ୭ should be pronounced as ଜୁଁ or ଜୁଁ by placing a dot (Nuqta) below them. ୫, ୬, ୭, ୮, and ୯ have a dot (Nuqta) below them to make them sound like ଖୁଁ, ଖୁଁ, ଗୁଁ, ଜୁଁ, ଫୁଁ. Two dots (double Nuqta) below the letter ୯ represent the letter ଜୁଁ. The reason for using the Kaithi letters above is to spell the Persian word correctly.

Name

The proposed character names are changeable as per Ad-Hoc decision.

Allocation

The proposed character should be encoded in the 'Kaithi' Block (U+11080 - U+110CF).

Character Data

Character Properties: `UnicodeData.txt`

`110C3 ; KAITHI LETTER ZHA;Lo;0;L;;;;N;;;;`

Indic Syllabic Category: `IndicSyllabicCategory.txt`

`110C3 ; Consonant # Lo [01] KAITHI LETTER ZHA`

Linebreaking Category: `lineBreak.txt`

`110C3 ; AL # Lo [01] KAITHI LETTER ZHA`

References

[1] Kishore, Naval. 1881. *କାଇଥି ବାରାନମାଳା ପାହିଲା ହିସାହା* [Kaithi Baranamala Pahila Hissaha [Kaithi Alphabets Part 1]]. Munshi Naval Kishore Press, Lucknow.

https://archive.org/details/Bczy_kaithi-varna-mala-part-1-1881-naval-kishore-press/mode/1up

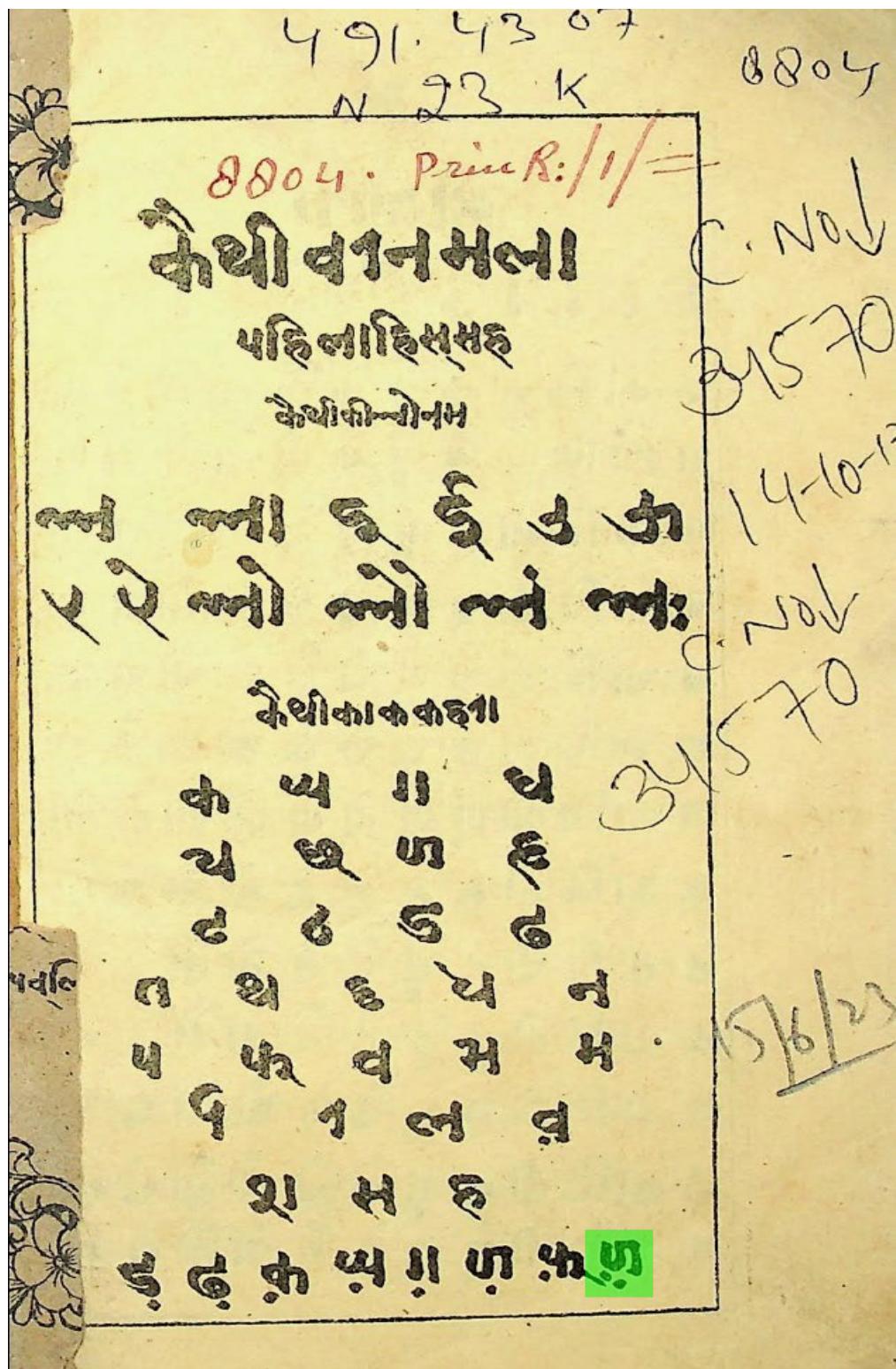


Fig 1: Kaithi alphabet chart. (Kishore 1881:1)

૪

ત તા તિ તી ત તુ તે તે તુ તો તુ તુ
 ત તા તિ તી ત તુ તે તે તુ તો તુ તુ
 તા તા તિ તી તુ તુ તે તે તુ તો તુ તુ
 તુ તુ તિ તી તુ તુ તે તે તુ તો તુ તુ
 તુ તુ તિ તી તુ તુ તે તે તુ તો તુ તુ
 તુ તુ તિ તી તુ તુ તે તે તુ તો તુ તુ
 તુ તુ તિ તી તુ તુ તે તે તુ તો તુ તુ

તુ તુ તિ તી તુ તુ તે તે તુ તો તુ તુ

Fig 2: Vowel sign positions with letter ZHA. (Kishore 1881:4)

૬

ય	યવા	ય યક્કિય
ય	યવા	ય યક્કિય

Fig 3: Word યાય written by letter ZHA. (Kishore 1881:7)

**PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹**

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from <http://std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.

Please ensure you are using the latest Form from <http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.
See also <http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

A. Administrative

1. Title:	Proposal to Encode the Kaithi Letter ZHA in Unicode	
2. Requester's name:	Biswajit Mandal (biswajitmandal.bm90@gmail.com)	
3. Requester type (Member body/Liaison/Individual contribution):	Individual Member	
4. Submission date:	16/12/2025	
5. Requester's reference (if applicable):		
6. Choose one of the following:		
This is a complete proposal:	Yes	
(or) More information will be provided later:	No	

B. Technical – General

1. Choose one of the following:		
a. This proposal is for a new script (set of characters):	Yes	
Proposed name of script:	Kaithi	
b. The proposal is for addition of character(s) to an existing block:		
Name of the existing block:		
2. Number of characters in proposal:	1	
3. Proposed category (select one from below - see section 2.2 of P&P document):		
A-Contemporary	<input checked="" type="checkbox"/>	B.1-Specialized (small collection)
C-Major extinct	<input type="checkbox"/>	D-Attested extinct
F-Archaic Hieroglyphic or Ideographic	<input type="checkbox"/>	E-Minor extinct
		G-Obscure or questionable usage symbols
4. Is a repertoire including character names provided?	Yes	
a. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document?	Yes	
b. Are the character shapes attached in a legible form suitable for review?	Yes	
5. Fonts related:		
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	Biswajit Mandal (biswajitmandal.bm90@gmail.com)	
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	Biswajit Mandal (biswajitmandal.bm90@gmail.com)	
6. References:		
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	Yes	
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	Yes	
7. Special encoding issues:		
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	
	See proposal for additional details	
8. Additional Information:		
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 Unicode Character Database (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.		

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?	<input type="checkbox"/> No
If YES explain	
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?	<input type="checkbox"/> Yes
If YES, with whom?	
Bhojpuri speaking community of India	
If YES, available relevant documents:	
See citations in text proposal	
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	<input type="checkbox"/> Yes
Reference: See text of proposal	
4. The context of use for the proposed characters (type of use; common or rare)	<input type="checkbox"/> Common
Reference: See text of proposal	
5. Are the proposed characters in current use by the user community?	<input type="checkbox"/> Yes
If YES, where? Reference: See text of proposal	
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	<input type="checkbox"/> No
If YES, is a rationale provided?	
If YES, reference:	
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	<input type="checkbox"/> N/A
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	<input type="checkbox"/> No
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?	<input type="checkbox"/> No
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character?	<input type="checkbox"/> Yes
If YES, is a rationale for its inclusion provided?	
If YES, reference: See text of proposal	
11. Does the proposal include use of combining characters and/or use of composite sequences?	<input type="checkbox"/> No
If YES, is a rationale for such use provided?	
If YES, reference:	
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?	
If YES, reference:	
12. Does the proposal contain characters with any special properties such as control function or similar semantics?	<input type="checkbox"/> No
If YES, describe in detail (include attachment if necessary)	

13. Does the proposal contain any Ideographic compatibility characters?	<input type="checkbox"/> No
If YES, are the equivalent corresponding unified ideographic characters identified?	
If YES, reference:	

UnicodeData.txt

```
1 110C3 ; KAITHI LETTER ZHA;Lo;0;L;;;;N;;;;;
```

IndicSyllabicCategory.txt

```
110C3 ; Consonant # Lo [01] KAITHI LETTER ZHA
```

LineBreak.txt

```
1 110C3 ; AL # Lo [01] KAITHI LETTER ZHA
```