



CENTRAL INSTITUTE OF INDIAN LANGUAGES

DEPARTMENT OF HIGHER EDUCATION

Ministry of Human Resource Development, Government of India

Manasagangotri, Mysore - 570 006, INDIA



Linguistic Data Consortium for Indian Languages (LDC-IL)

LDC-IL Transliteration Scheme Version 0.1A

On the basis of future comments, suggestions, necessities etc, more symbols may be added.

The term "Unicode" stands for the Unicode Standard 5.0

This listing does not imply any sorting order.

I	II	III	IV
01	अ	a	a
02	आ	ā	A
03	इ	i	i
04	ई	ī	I
05	उ	u	u
06	ऊ	ū	U
07	ऋ	ṛ	x
08	ॠ	ṝ	X
09	ऌ	ḷ	q
10	ॡ	ḹ	Q
11	ऐ	e	e
12	ए	ē	E
13	ऐ	ai	ai
14	ऎ	ê	ae
15	ओ	o	o
16	ओ	ō	O
17	औ	au	au
18	ऑ	ô	ao
19	अं		aM
20	अँ	m̃	m'
21	अः		aH
22	क	ka	ka

23	ख	kha	kha
24	ग	ga	ga
25	घ	gha	gha
26	ङ	ṅa	ng'a
27	च	ca	ca
28	छ	cha	cha
29	ज	ja	ja
30	झ	jha	jha
31	ञ	ña	nj'a
32	ट	ṭa	Ta
33	ठ	ṭha	Tha
34	ड	ḍa	Da
35	ढ	ḍha	Dha
36	ण	ṇa	Na
37	त	ta	ta
38	थ	tha	tha
39	द	da	da
40	ध	dha	dha
41	न	na	na
42	न	na	n'a
43	प	pa	pa
44	फ	pha	pha
45	ब	ba	ba
46	भ	bha	bha
47	म	ma	ma
48	य	ya	ya
49	य	yā	Ya
50	र	ra	ra
51	र	ra	Ra
52	र	ra	r'a
53	ल	la	la
54	ळ	ḷa	La

55	ळ	za	Za
56	व	va	va
57	श	śa	sha
58	ष	ṣa	Sa
59	स	sa	sa
60	ह	ha	ha
61	ग		g"a
62	ज		j"a
63	ङ		D"a
64	ब		b"a
65	क		k'a
66	ख		kh'a
67	ग		g'a
68	ज		j'a
69	ङ		D'a
70	ढ		Dh'a
71	फ		ph'a
72	१		\1
73	२		\2
74	३		\3
75	४		\4
76	५		\5
77	६		\6
78	७		\7
79	८		\8
80	९		\9
81	०		\0
82			\.
83			\..
84	s		\s
85	च		c'a
86	छ		ch'a
87	'		"

88	ॐ	@M
89	'	\'
90	"	\"
91	ٹ	s`a
92	ذ	za
93	ص	s'a
94	ض	z'a
95	ط	t`a
96	ظ	z`a
97	ع	A'
99	ھ	Ha
100	ء	a'

=====

Note AA: Columns I to IV stand for:

I=Serial Number (this is the number to be used in further discussions)

II= Devanagari Character used for illustrative purpose. The corresponding characters for other Indic scripts have to be derived by using the appropriate xml tagging and interpretation.

III= ISCII input equivalent (BIS standard ISCII-91 = IS13194:1991)

IV= ASCII keyboard keystroke(s) that will call the corresponding Unicode codepoints

=====

NOTE AB: 11, 15 stand for the short half-open vowels (front and back respectively) found especially in Dravidian languages.

NOTE AC: 14, 18 stand for the open vowels (front and back respectively) specifically marked in Marathi orthography and to some extent in Hindi orthography.

NOTE AD: 42 stands for Tamil ண

NOTE AE: 49 stands for Bangla and Asamiya ঞ

NOTE AF: 51 stands for Tamil ற Kannada ಱ Telugu ఱ Malayalam ౠ

NOTE AG: 52 stands for the so-called 'eyelash r' used in Marathi and Nepali.

NOTE AH: 54 stands for Tamil ள, Malayalam ూ Telugu ళ Kannada ಳ

NOTE AI: 55 stands for Tamil ழ, Malayalam ృ

NOTE AJ: 61, 61, 63, 64 stand for the underlined ग ज ङ ब respectively which are implemented for Sindhi implosives (these are the Unicode U+ 097B, 097C, 097E, 097F respectively)

NOTE AK: 65-71 are the Devanagari characters with nukta (listed as U+ 0958 to 095E respectively)

NOTE AL: 72-81 are the Devanagari digits (U+0966 to +0970)

NOTE AM: 82 and 83 are the Devanagari Danda and Double Danda respectively (U+0964, 0965 respectively)

NOTE AN: 84 is the Devanagari avagraha character (U+093D)

NOTE AO: 85-86 are the dental affricates of Kashmiri (85 stands for च with nukta and 86 stands for छ with nukta)

NOTE AP: 87 is the {sur} symbol of Dogri

NOTE AQ: 88 is the ॐ symbol of Devanagari (this is a single character in Unicode and is different from the string ओम् or the string ओम)

NOTE AR: 89 and 90 stand for the ASCII 'single quote' and "double quote" respectively (thus 87 stands for the Dogri {sur} symbol whereas 89 stands for the ASCII 'single quote')

NOTE AS: 91 stands for Urdu ث.

NOTE AT: 92 stands for Urdu ذ

NOTE AU: 93-94 stands for Urdu ض and ص

NOTE AV: 95-96 stands for Urdu ط and ظ

NOTE AW: 97 stands for Urdu ع

NOTE AX: 98 stands for Urdu ه

NOTE AY: 99 STANDS FOR Urdu ء

**Sample Scheme of converting
LDC-IL {Base XML string} into the corresponding {Unicode5.0} string
An Illustration**

Column A gives an illustrative {XML string in ASCII}. All the data in Indic scripts will be of this structure.

Column B gives the desired corresponding processed {Unicode5.0 output string}

	Output in Script:
<SDvn>bhAratamAtA</SDvn>	भारतमाता
<SKan>bhAratamAtA</SKan>	ಭಾರತಮಾತಾ
<SGjr>bhAratamAtA</SGjr>	ભારતમાતા
<SBgl>bhAratamAtA</SBgl>	ভারতমাতা
<SGrm>bhAratamAtA</SGrm>	ਭਾਰਤਮਾਤਾ
<STlg>bhAratamAta</STlg>	ಭಾರತಮಾತ
	Output in Phonemic Transcription:
<HndPhnm>bhAratamAtA</HndPhnm >	ḅa:ratma:ta:
<PnjPhnm>bhAratamAtA</ PnjPhnm >	ḅa:ratma:ta:
	Output in Phonetic Transcription:
<HndPhnt>bhAratamAtA</ HndPhnt >	ḅa:ratma:ta:
<PnjPhnt>bhAratamAtA</ PnjPhnt >	pá:ratma:ta:

Note: The above xml tags for languages are just for illustration. A comprehensive language name tag set along with other tag sets (that form part of the Global Tag Set of LDC-IL) will be made available.