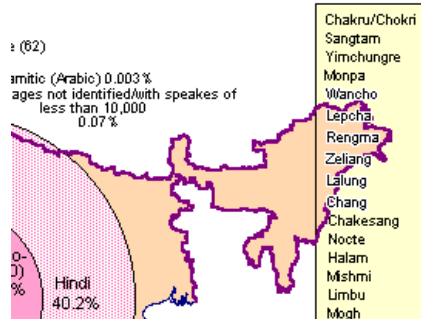


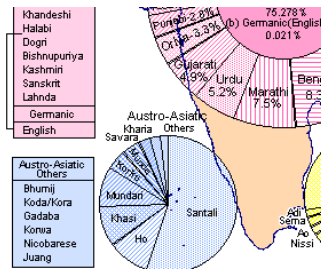
FIRST FACE-TO-FACE MEETING OF
THE NEO-BRAHMI SCRIPT
GENERATION PANEL; 10 am-5 pm

16 July, 2015
COURTYARD BY
MARRIOTT, Pune

The Internet Corporation for Assigned Names and Numbers An ICANN Event



Multiscriptal India Diversity on a Unified Platform



Udaya Narayana Singh
VISVA-BHARATI, SANTINIKETAN

E-mail: unscil@yahoo.com

Chair, Centre for Endangered Languages
& Professor, Rabindra Bhavana

Centre for Development of Advanced Computing, Pune

सी डैक
CDAC
Gist

What is in a Name?

- ❑ At the outset, we must thank ICANN for bringing so many of us – members of a variegated writing tradition – on board.
- ❑ The challenges & opportunities for research would soon be clear once we begin looking for answers to the proverbial question (What is in a name?).
- ❑ But frankly, speaking in a lighter vein, as we begin our journey together, I was wondering – specialists of your own scripts, as to why am I here, or how could I be useful to the endeavor?
- ❑ The namespaces on the internet must give us a stable and secure operation once a new set of writing systems are allowed entry into the domain name registry, or else this great effort to democratize the web would suffer.
- ❑ Thus ICANN would promote as well as regulate use of multicriptal domain names where we must all help this not for profit endeavor.
- ❑ As for the Neo-Brahmi situation, let's first look at the big picture.



A 4th-5th Century Pallava Inscription
from Bujong Valley, Malaysia

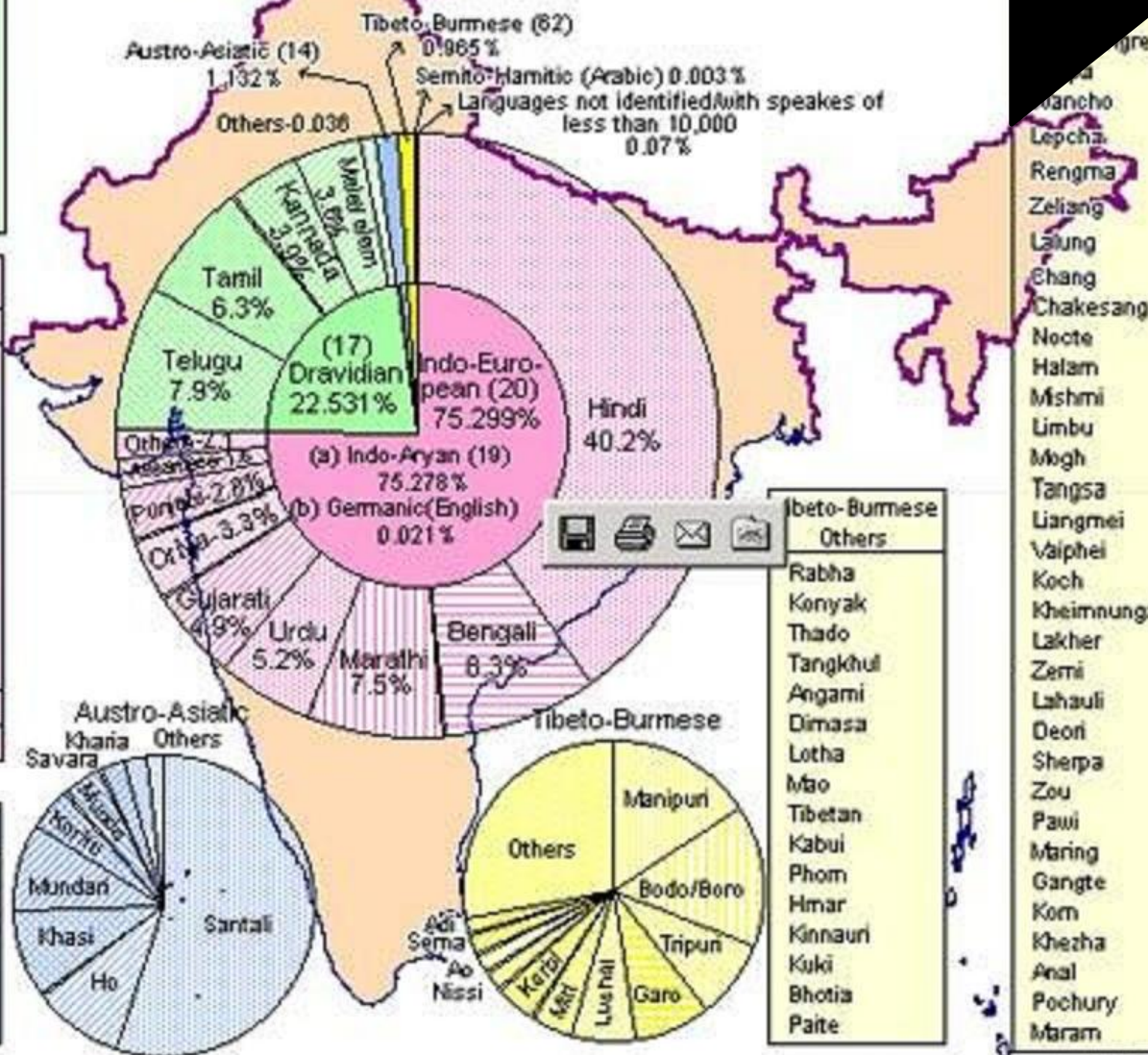
INDIA Indian Language Families (Based on the 1991 Census)

**The Big Picture
- 1**

Dravidian Others
Gondi
Tulu
Kurukh/Oraon
Kui
Koya
Khond/Kondh
Kisan
Malto
Kolami
Coorgi/Kodagu
Parji
Jatapu
Konda

Indo-European Others
Indo-Aryan
Bhili/Bhilodi
Sindhi
Nepali
Konkani
Khandeshi
Halabi
Dogri
Bishnupuriya
Kashmiri
Sanskrit
Lahnda
Germanic
English

Austro-Asiatic Others
Bhumij
Koda/Kora
Gadaba
Korwa
Nicobarese
Juang



- ❑ **1576** rationalized mother-tongues (MTs) as per Census 2001
- ❑ **122** languages with 10,000 plus speakers;
- ❑ **146** used in radio network now; Before 1939, English, Bangla, Hindustani, & later Telugu, Tamil, Marathi, Gujarati were aired.
- ❑ **35** of them publish 3954 newspapers as in 1971 (doubled in '03, with Hindi accounting for 2507, Urdu 534 & English 407.
- ❑ **22** Constitutional languages;
- ❑ **29** have one million+ speakers (E.g Bhili, Gondi, Khasi, Tulu, Mundari, Ho etc)
- ❑ **24** have highest/ Akademi awards

The Big Picture- 2

ROM	DEV	GUJ	PUN	BEN	ORI	TEL	KAN	TAM	MAL	SINH	URD	SIND	SAN	MAN
k	क	ક	ਕ	କ	କ	క	ಕ	க	ക	ක	ک	ڪ	କ	𑖅
kh	ख	ખ	ਖ	খ	ଖ	ఖ	ಖ	—	ഖ	ඬ	کھ	ڪھ	କ୍	𑖆
g	ग	ગ	ਗ	গ	ଗ	గ	ಗ	—	ഗ	ග	گ	گ	ଗ	𑖇

- **Four Major Language Families & a few minor ones**
- **14 major writing systems in use but 66 scripts in all.**

- ❑ **1796** other Mother-Tongues not yet classified – MTSI and SPPEL Projects to reveal more details;
- ❑ **Large variations** : Over 41% speak Hindi with 49 varieties & 422 million speakers to Maram (Manipur) with 10,144 speakers;
- ❑ **Large non-scheduled langs**; eg Bhili - 9.58 million speakers;
- ❑ **97.7%** speak only 20-odd Indo-Aryan & Dravidian languages
- ❑ **2.32%** of Indian population speak abt 97% of other languages

Brâhmî & Kharoṣṭhî

- Georg Bühler (1895) argued that Brâhmî (from the 8th cent BC) is the mother script to numerous writing systems [*On the Origin of the Indian Brahma Alphabet*]
- David Diringer (1953: *The Alphabet*) questioned this date, and thought the origin was perhaps Aramaic.
- Hans Jensen (1935) relied on Buhler.
- The thesis that Brâhmî was developed out of a Semitic prototype in pre-Mauryan India (before 7th Cent BC?) was accepted by the western scholars but rejected by South Asians, who derive it from Indus Valley script (3300-1300 BC)

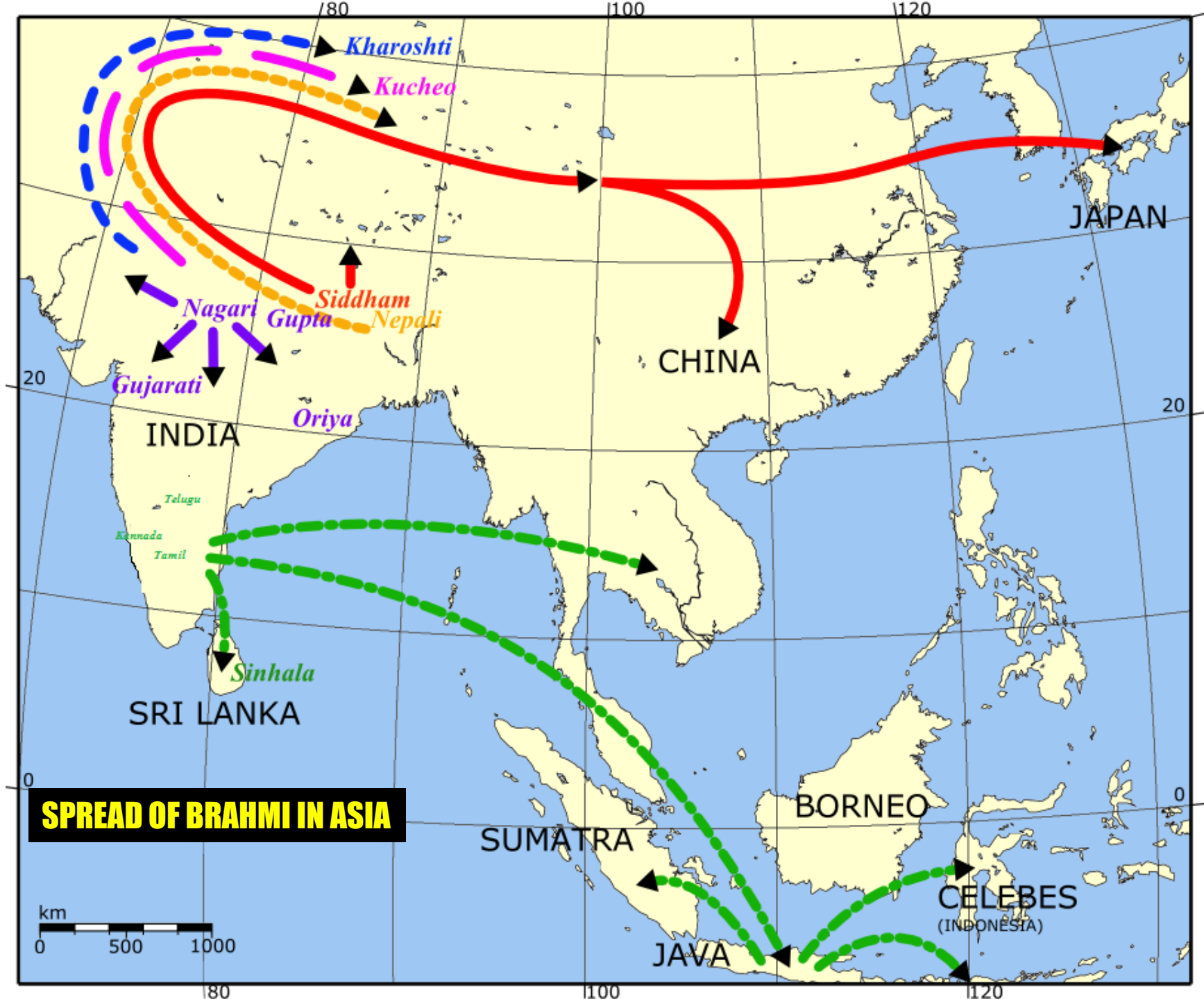
𑀓 a	𑀕 ā	𑀖 i	𑀗 ī	𑀘 u	𑀙 ū
𑀚 e	𑀛 ai	𑀜 o			
𑀞 ka	𑀟 kha	𑀠 ga	𑀡 gha	𑀢 ṇa	
𑀣 ca	𑀤 cha	𑀥 ja	𑀦 jha	𑀧 ṇa	
𑀨 ṭa	𑀩 ṭha	𑀪 ḍa	𑀫 ḍha	𑀬 ṇa	
𑀭 ta	𑀮 tha	𑀯 da	𑀰 dha	𑀱 na	
𑀲 pa	𑀳 pha	𑀴 ba	𑀵 bha	𑀶 ma	
𑀷 ya	𑀸 ra	𑀹 la	𑀺 ḷa	𑀻 va	
𑀼 śa	𑀽 ṣa	𑀾 sa	𑀿 ha		

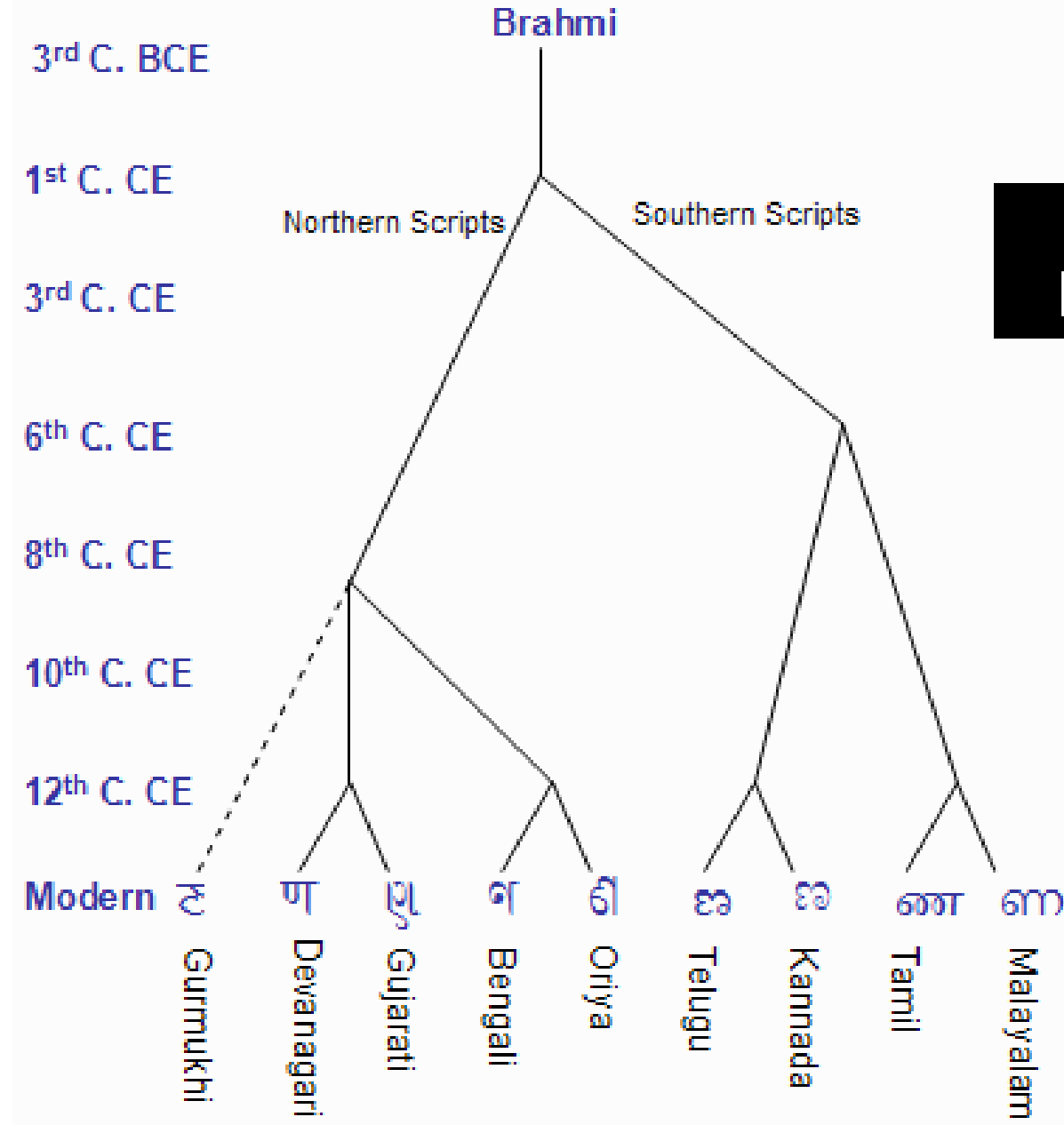
Brâhmî

𑀓 a	𑀕 ā	𑀖 i	𑀗 ī	𑀘 u	𑀙 ū
𑀚 e	𑀛 ai	𑀜 o			
𑀞 ka	𑀟 kha	𑀠 ga	𑀡 gha	𑀢 ṇa	

Kharoṣṭhî

A greater clarity on the Indus valley script & discovery of Greek and Aramaic inscriptions of Ashokan times from Pakistan/ Afghanistan, and the studies by Fussman, von Hinüber, and Falk show that Brâhmî, & Kharoṣṭhî emerged in the Mauryan period (321-232 BC).





**An Early
Divergence**

The diagram shows an early divergence between North and South Indian scripts as shown by Peter T. Daniels & William Bright (1996) *The World's Writing Systems*, OUP.

Multi-scriptality - a reality!

One Script for Many Languages

- Devanagari used by official languages : Hindi, Marathi, Konkani, Nepali, Sanskrit, Sindhi, Dogri and Santhali are written, not to mention around 30 more languages
- Eastern Brahmi Base (close to Bengali) has been used by Assamese, Bangla, Bodo, Manipuri and Maithili.
- Modi was used primarily by Marathi until 19th century and also by Urdu, Kannada, Gujarati Hindi and Tamil.

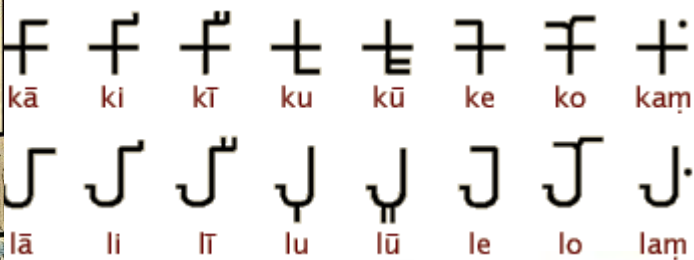
ॐ नमो भगवते वासुदेवाय ।
ऐह प्रतीतिदा ।
नमो नमो स्वर्गदेवा ।
पद्मरुपा ॥






















Modi Script sample

झ छ ग घ ङ
च छ ज झ
ठ ड उ द ण
त थ द ध न
प फ ब म य
र र २ ४ ए ऌ
ष ष ण क ङ ह

Many Scripts for One Language

In contrast, there could also be languages like **Konkani** written in Roman, Devanagari, Kannada, and Malayalam scripts. Earlier **Santali** written in Roman, Ol Chiki, Devanagari, Oriya and Bangla scripts. Or, **Bodo** in Devanagari, Roman and Assamese.



		
a	ā	i
		
e	ai	o
		
ka	kha	ga
		
ca	cha	ja
		
ṭa	ṭha	ḍa
		
ta	tha	da
		
pa	pha	ba
ya	ra	la

NEW SCRIPTS EMERGING AS ALWAYS

	a	i	u	e	o	Meitei Meyek					
ka	ka	kha	ga	gha	na	𑜀𑜂𑜆𑜇	𑜀𑜃𑜆𑜇	𑜀𑜆𑜂𑜆𑜇	𑜀𑜆𑜃𑜆𑜇	𑜀𑜆𑜄𑜆𑜇	
ca	ca	cha	ja	jha	ña	𑜀𑜂𑜆𑜇	𑜀𑜃𑜆𑜇	𑜀𑜆𑜂𑜆𑜇	𑜀𑜆𑜃𑜆𑜇	𑜀𑜆𑜄𑜆𑜇	
ta	ta	tha	da	dha	ña	𑜀𑜂𑜆𑜇	𑜀𑜃𑜆𑜇	𑜀𑜆𑜂𑜆𑜇	𑜀𑜆𑜃𑜆𑜇	𑜀𑜆𑜄𑜆𑜇	
pa	pa	pha				𑜀𑜂𑜆𑜇	𑜀𑜃𑜆𑜇	𑜀𑜆𑜂𑜆𑜇	𑜀𑜆𑜃𑜆𑜇	𑜀𑜆𑜄𑜆𑜇	
ya	ya	ra	la			𑜀𑜂𑜆𑜇	𑜀𑜃𑜆𑜇	𑜀𑜆𑜂𑜆𑜇	𑜀𑜆𑜃𑜆𑜇	𑜀𑜆𑜄𑜆𑜇	
sa	sa	sa	sa	ha		𑜀𑜂𑜆𑜇	𑜀𑜃𑜆𑜇	𑜀𑜆𑜂𑜆𑜇	𑜀𑜆𑜃𑜆𑜇	𑜀𑜆𑜄𑜆𑜇	

Saurashtri Lipi


Kharosthi



ka	kha	ga	gha	ña
ca	cha	ja	jha	ña
ṭa	ṭha	ḍa	ḍha	ṇa
ṁ	ṭh	ḍ	ḍh	ṇ
ta	tha	da	dha	na
pa	pha	ba	bha	ma

Akkhara Muni (Pali)






ka ku ki ké ko kau keu/ke



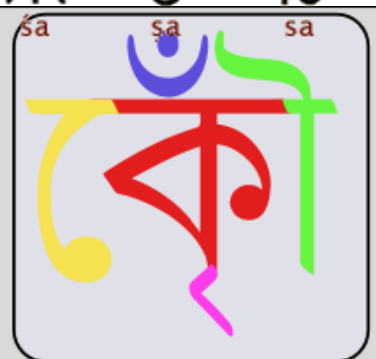
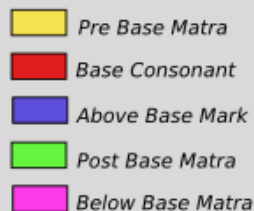






kēa/kah ka'/kak kai kar kang kan -k

Rejang<Kawi<Brahmi



अ आ इ ई उ ऊ
ए ऐ ओ औ अं अः
क ख ग घ ङ
च छ ज झ ञ
ट ठ ड ढ ण
त थ द ध न

Nagari Lipi

Challenges Due To Internal Complications –

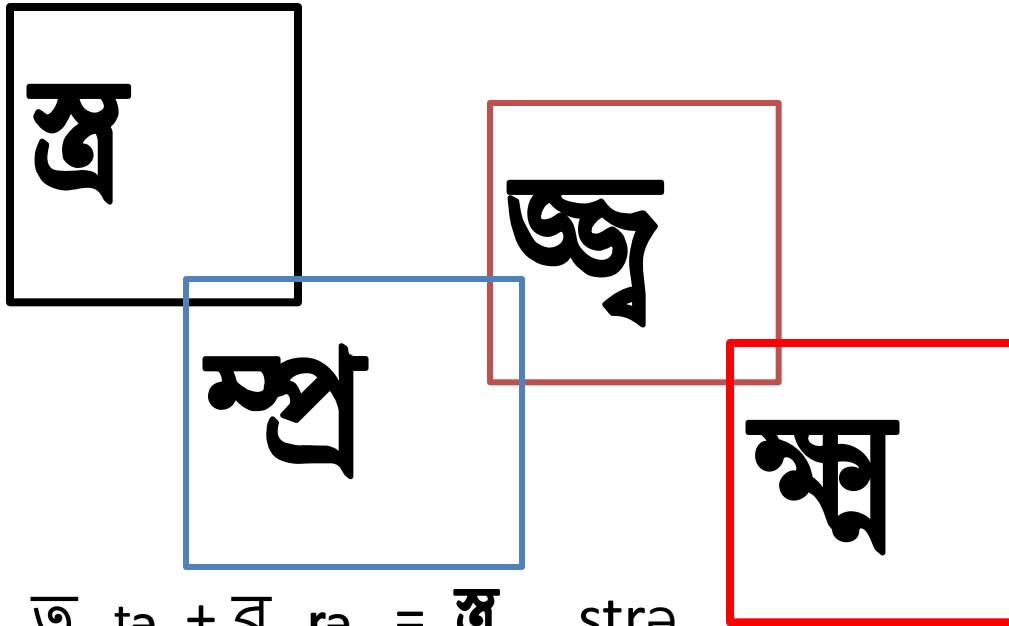
e.g. অংক vs অঙ্ক vs অঙক

- The variety of conjuncts and the accent markings in old script-variants of **Brâhmi**, known as **Vedic**, **Sâmavedic**, and **Grantha** make these heritage scripts a challenge.
- Each modern Brâhmi system has numerous glyphs as well as conjoined or combined letters. To take an example, Telugu alone has 720 glyphs, Bengali-Assamese 520, Meitei-Meyek 509, Gujarati 436, Devanagari & Malayalam 425, Oriya 410, Kannada 399, Punjabi 358, and Tamil 316.
- Bodo, drawing largely on Devanagari has 678 glyphs. Santali Ol Chiki, a newly developed script also has 678 glyphs, and so had the older script of Maithili – known as Tirahuta, or Dogri using 678 glyphs.
- There are a total 1520 glyphs are in the **Vedic Sanskrit Open Type Font**, whereas Sâmavedic has 1489 glyphs, and Grantha 1143 glyphs.
- Compare it with Urdu: Nataaliq 563, Kashmiri 584 & Sindhi 411 glyphs.

এত	[æ̣to] "so much",
এ্যাকাডেমী	[æ̣kademi] "academy",
অ্যামিবা	[æ̣miba] "amoeba"

Kind of Problems at Hand

A COMPARATIVE PICTURE

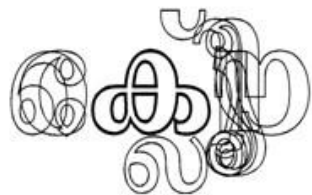


স sə + ত তə + র rə = স্ত্র strə,

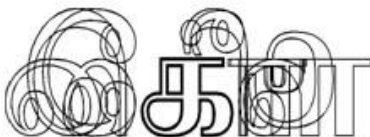
ম mə + প pə + র rə = ম্প্র mprə,

জ jə + জ jə + র wə = জ্জ jjwə,

ক kə + ষ ṣə + ম mə = ক্ষ্ম kṣmə



Malayalam



Tamil



Gujarati



Devanagari



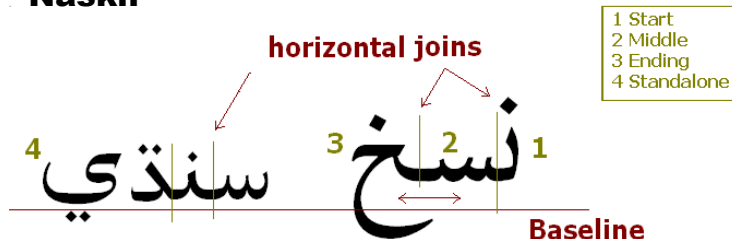
Oriya

Complications

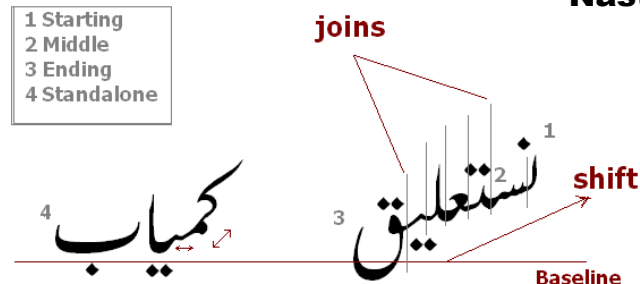
Irrespective of the South-North division, each of these writing systems has numerous difficulties for the font developers because of the way their syllabic structures are captured by the graphics – protruding below, above, and on the aisles as well as hanging letters.

Script Grammar for all Indic systems to capture their aesthetics.
Only Sindhi uses system – **Naskh** (linear-seamless) **but** Kashmiri & Urdu use **Nastaliq** (Right-to-left plus Top-to-bottom):

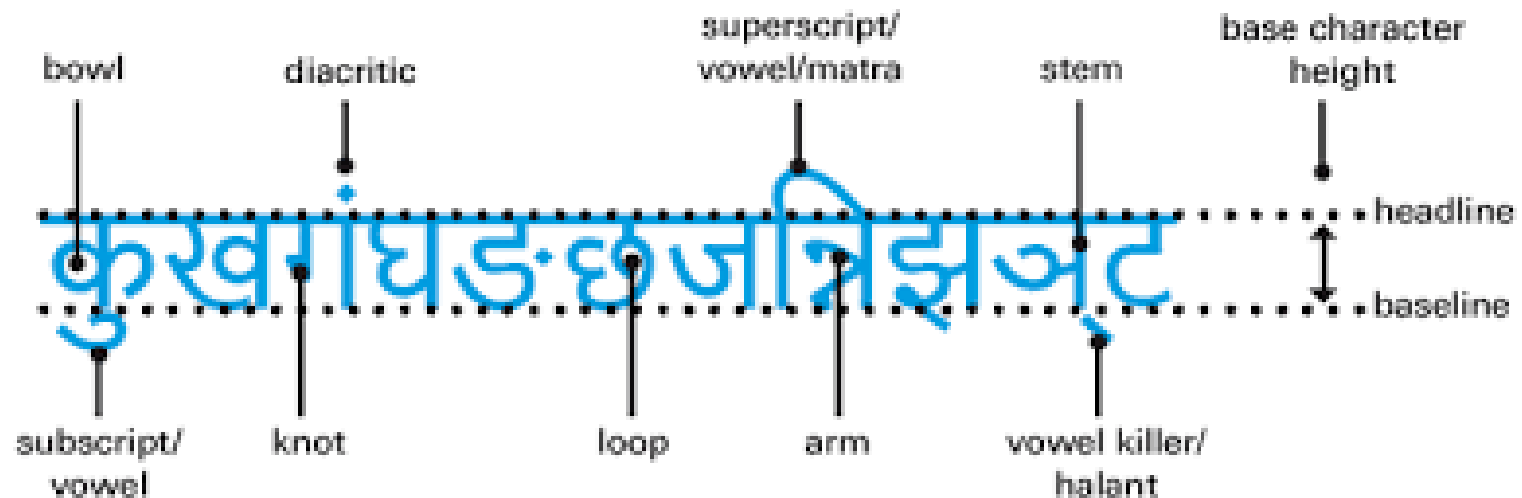
Naskh



Nastaaliq



- ❑ There are many more issues that would be spelt out today in the next few presentations .
- ❑ All these need resolution if Indian Scripts must be used for Domain Name Registration
- ❑ Let's volunteer to help ICANN resolve these issues for Indians & South/South-East Asians.



Thank you