NEO-BRĀHMĪ

AN OVERVIEW

Neo-Brāhmī Generation Panel
16 July 2015

Agenda

- Setting things in place
- Why the name Neo-Brāhmī
- The origins of Brāhmī.
- Features of Brāhmī.
- Basic Evolution of Brāhmī
- Principal Neo-Brāhmī Languages.
- Target Scripts and justification of the choice
- Neo-Brāhmī LGR is Pan-Asian and not Pan-Indian

Setting things in place

AIM:

Provide a broad sweeping overview of the Neo-Brāhmī LGR and answer to the question:

What is expected of me?

FLOW: CONCENTRIC APPROACH

Overview of Brāhmī

Unicode: A Primer

IDN: An Overview

Background of the IDN Variant TLD's Program

Label Generation Rule-set Process for the Root Zone

Project Timelines

Why the name Neo-Brāhmī ?

Background:

- ICANN has identified six generation panels, all of which are based on writing systems. These are: Latin, Greek, Cyrillic, CJK, Arabic and Indic. Other minor panels such as Armenian have also been approved and many more are in the pipe-line
- For each group, Label Generation Rules have to be developed
- In the case of the Indic family, we have named our group: Neo- Brāhmī
- The Brāhmī script has been the progenitor of all scripts used to write Modern Indo-Aryan languages, Dravidian and to a lesser extent scripts of the Tibeto-Burman and Munda families. It was also adopted by a large number of cultures in Southeast Asia to transcribe their languages: Burmese, Thai, Lao, Khmer (in South-East Asia), and others in Central Asia
- The Neo-Brāhmī group is so named to cover all such scripts used today and which are based on Brāhmī

The origins of Brāhmī.

Brāhmī evolved around the 3rd Century BC. The origin of Brāhmī is a debatable question. Some scholars treat it as a based on the Semitic writing system: Phoenician or Aramaic. Others view it as an indigenous Indian invention, often associated with the Indus Valley script.

http://www.ancientscripts.com/brahmi.html

Features of Brāhmī

- Brāhmī is written from left to right
- It has an angular shape. As it evolved this angular feature was gradually replaced by rounded shapes in cultures where palm leaves where used as a medium of written communication.
- The main feature of Brāhmī is the written syllable or akṣara a concept admitting a full Consonant or Vowel as a node.
- Vowels admit Vowel Modifiers such as nasals or vowel lengtheners.
- Consonants are at times modified by a combining mark functioning as "vowel-killer" (termed Halanta), truncating the following vowel. These constitute conjuncts
- In turn these can be modified by vowel signs: Matras and further by Nasals or Vowel lengtheners
- The adjuncts to the Vowel or Consonant nodes are appended in a strict rule-order.
- This feature has been remarkably stable over the evolution of Brāhmī and has been followed by all the later Indic and Southeast scripts derived from the script.

Basic Evolution of Brāhmī.

- Around the 3rd Century AD, Brāhmī script in India had already divided itself into two main styles commonly termed as Northern and Southern.
- Roughly the Northern branch was adopted by the Indo-Aryan and Tibeto-Burman family and to a lesser extent by the Dardic.
- The Southern family coincides with the Dravidian family and which in turn spread out to South-East Asia.
- The distinct forms of modern scripts of India derived from Brāhmī took a stable form between the tenth and fifteenth centuries.

Basic Evolution of Brāhmī Northern Branch

- The Northern Branch of Brāhmī gave rise to what is termed as the Indo-Aryan family.
- Bengali and Odia constituted one branch,
- Devanāgarī (which later split into Gujarati and Devanāgarī script), constituted the other.
- These developed in succeeding centuries into sub-varieties which can be broadly divided into Western and Eastern varieties in the north and into Deccan and Peninsular varieties in the south.
- Other branches of this family include the scripts of the western Himalayan regions, of which the modern Gurumukhi or Punjabi script is the only important modern survivor. The Śāradā script of Kashmir also belongs to this group.

Basic Evolution of Brāhmī: Southern Branch.

- The Southern Branch is characterized by the Dravidian family.
- Although the evolution itself started around the 3rd Century C.E., by the 12th Century the Southern branch of Brāhmī had already split into smaller script families: Two large families evolved which one gave rise to Telugu and Kannada and the other to Tamil and Malayalam.
- The Sinhala script was imported from North India but was influenced at various stages of its development by south Indian scripts.

Basic Evolution of Brāhmī: A Time-Machine

http://www.ancientscripts.com/brahmi.html

The World's Writing System, edited by Peter T. Daniels and William Bright, New York - Oxford, Oxford University Press, 1996,

Principal Neo-Brāhmī Languages

© Britle

Principal Neo-Brāhmī Languages

- Due to commercial, societal and religious factors the Brāhmī script spread rapidly through Central, South Asia and South-East Asia giving rise to three major branches of the script.
- The Central Asian influenced the writing system of Tibetan and the Himalayan Languages.
- The South Asian Branch roughly covers the Indo-Aryan and Dravidian languages including Sinhala
- The South-East Asian branch includes Lao, Burmese, Shan and a large number of languages of South-Eastern Asia.

Principal Neo-Brāhmī Languages

Central Asian Scripts
Manichaean
Mongolian
Old Turkic
Phags-Pa
Tibetan

Southeast Asian Scripts
Cham
Kayah Li
Khmer
Khmer Symbols
Lao
Myanmar
Myanmar Extended-A
Myanmar Extended-B
New Tai Lue
Pahawh Hmong
Pau Cin Hau
Tai Le
Tai Tham
Tai Viet
Thai

South Asian Scripts
Ahom
Bengali and Assamese
Brahmi
Chakma
Devanagari
Devanagari Extended
Grantha
Gujarati
Gurmukhi
Kaithi
Kannada
Kharoshthi
Khojki
Khudawadi
Lepcha
Limbu
Mahajani
Malayalam
Meetei Mayek
Meetei Mayek Extensions
Modi
Mro
Multani
Ol Chiki
Oriya (Odia)
Saurashtra
Sharada
Siddham
Sinhala
Sinhala Archaic Numbers
Sora Sompeng
Syloti Nagri
Takri
Tamil
Telugu
Thaana
Tirhuta
Vedic Extensions
Warang Citi

Excerpted from Unicode: http://unicode.org/charts/

Principal Neo-Brāhmī Languages Central Asian Scripts.

Brāhmī influenced the writing systems of the Himalayan region and *Tibetan* and *Dzongkha* are among the languages of the Himalayan region whose scripts are based on Brāhmī.

ALPHABETICAL PLAN OF THE TIBETAN LANGUAGE,

The five vowels: དབྱངས་ལྔ (*yāng ṇga*).

ཨ་ཨི་ཨུ་ཨེ་ཨོ། *a, i, u, e, o.*

The four vowel signs that are attached to the basic letter ཨ are called *gi-gü, shabkyü, ḍeng-bü* and *nāro*:

ཿ ི ུ ེ ོ། *i, u, e, o.*

The thirty consonants: གསལ་བྱེད་སུམ་ཅུ་རྩ (*sal-je süm-chü*).

ཀ ཁ ག ང།	ཚ ཚ ཛ ཉ།	ཏ ཐ ད ན།
ka, kha, ga, ña.	*ca, cha, ja, ña.*	*ta, tha, da, na.*
པ ཕ ཨ ཨ ཨ ཨ	ཙ ཚ ཛ ཝ	ཤ ཟ ཧ ཡ
pa, pha, ba, ma.	*tsa, tsha, dsa, wa.*	*sha, za, ha, ya.*
ར ལ ཤ ས	ཧ ཨ	
ra, la, ça, sa.	*ha, a.*	

ཀ rka	ག rga	ང rña	ཇ rja	ཉ rña	ཏ rta	ད rda	ན rna	པ rba			
ཁ rma	ཚ rtsa	ཛ rdza	ཀྱ rkya	ཁྱ rgya	ཁ rmya	ཀླ lka	ཁླ lga	ཀླ lña			
ཁ ltsa	ཁ lza	ཁ lta	ཁ lda	ཁ lpa	ཁ lba	ཁ lha	ཁ kla	ཁ gla			
ཁ bla	ཁ zla	ཁ rla	ཁ sla	ཁ kwa	ཁ khwa	ཁ gwa	ཁ cwa	ཁ ñwa			
ཁ twa	ཁ dwa	ཁ tshwa	ཁ žwa	ཁ zwa	ཁ rwa	ཁ lwa	ཁ šwa	ཁ swa			
ཁ ska	ཁ sga	ཁ sña	ཁ sña	ཁ sta	ཁ sda	ཁ sna	ཁ spa	ཁ sba			
ཁ sma	ཁ stsa	ཁ skya	ཁ sgya	ཁ sñya	ཁ sbya	ཁ snya	ཁ kya	ཁ khya			
ཁ gya	ཁ pya	ཁ phya	ཁ bya	ཁ mya	ཁ kṣa	ཁ gha	ཁ dha	ཁ bha			
ཁ dzha	ཁ dma	ཁ kra	ཁ khra	ཁ gra	ཁ tra	ཁ thra	ཁ dra	ཁ pra			
ཁ phra	ཁ bra	ཁ mra	ཁ šra	ཁ sra	ཁ hra	ཁ grwa	ཁ rgwa	ཁ rcwa			

© Dorjes Hugden © silcom.com

Principal Neo-Brāhmī Languages: South-east Asian Scripts.

The spread of Brāhmī in Southeast Asia originates from the Southern family, especially the influence of the Pallava script named after the dynasty of the Pallavas. By the late eighth century; the scripts began to acquire separate identities.
Brāhmī-based scripts are used to write languages of several different families in South-East Asia: Mon (Burma), Khmer (of Cambodia); Thai, Lao, Shan; Karen (Burma) and Burmese Javanese and other languages of Indonesia and the Philippines. Some of the major South-East languages using these scripts belong to the Austronesian (Jawi, Cham), Tai (Thai) and Tibeto-Burman (Karen, Mon, Shan, Burmese) families.

ก	ข	ค	ฆ	ง	จ	ฉ	ช	ซ	ฌ	ญ	ฎ
g	k	k	k	ng	j	ch	ch	s	ch	y	d
ฏ	ฐ	ฑ	ฒ	ณ	ด	ต	ถ	ท	ธ	น	บ
dt	t	t	t	n	d	dt	t	t	t	n	b
ป	ผ	ฝ	พ	ฟ	ภ	ม	ย	ร	ฤ	ฤา	ล
bp	p	f	p	f	p	m	y	r	reu	reu	l
ฦ	ฦา	ว	ศ	ษ	ส	ห	ฬ	อ		ฮ	
leu	leu	w	s	s	s	h	l	not sounded		h	

© quora.com
© www.opengroup.com

ในช่วงระยะแค่ 5 ปี มีนักท่องเที่ยวเพิ่มขึ้นเรื่อย ๆ ในปี 2534 มีตัวเลขนักท่องเที่ยวประมาณ 220.000 คน ทำกำไรให้กับมณฑลถึง 63 ล้านเหรียญสหรัฐ

อาจจะเป็นความใฝ่ฝันของคนคุนหมิงอยู่พอสมควรว่า เมื่อกรุงปักกิ่งเป็นตลาดท่องเที่ยวของอเมริกาและยุโรป คุนหมิงก็น่าจะเป็นตลาดท่องเที่ยวของชาวเอเชียตะวันออกและตะวันออกเฉียงใต้

Principal Neo-Brāhmī Languages South-east Asian Scripts

The figure below shows the evolution of the akṣara "ta" in major Southeast scripts influenced by Brāhmī.

The World's Writing System, edited by Peter T. Daniels and William Bright, New York - Oxford, Oxford University Press, 1996

Principal Neo-Brāhmī Languages South-east Asian Scripts

These roughly cover the scripts of India and Sri Lanka. They constitute by far the largest families derived from Brāhmī .

As a simplification it could be said that the Northern Branch covers the major Indo-Aryan script families and the Southern Branch coincides with the Dravidian.

Principal Neo-Brāhmī Languages South Asian Scripts.

- *Devanāgarī*: Devanāgarī is currently used for 11 out of 22 official languages of India (Boro/Bodo, Dogri, Hindi, Kashmiri, Konkani, Maithili, Marathi, Nepali, Sanskrit, Santhali and Sindhi) and around 45 other languages especially the related Indo-Aryan languages: Bagheli, Bhili, Bhojpuri, Himachali dialects, Magahi, Newari and Rajasthani and its dialects: Marwari, Mewati, Shekhawati, Bagri, Dhundhari, Harauti and Wagri. The script is also used in Fiji to represent Fiji Hindi. Hindi is also used in Mauritius, Malaysia, England, Canada, South Africa, Indonesia as well as emigrant communities around the world.

- *Gujarati*: Used for writing Gujarati and Kacchi, Gujarati is extensively spoken in large parts of Africa, Madagascar, UK and the USA as well as by emigrant communities around the world.

- *Gurumukhi* which evolved separately in the Northern family is used to write the Punjabi language in the Indian state of Punjab and elsewhere in India. Gurmukhi stabilised around the 16th century when it was used to transcribe the holy Granth Sahib.

- *Bengali:* Often termed as Bangla by linguists and grammarians is historically related and similar in design to the Devanāgarī script and with one or two exceptions has the same consonant and vowel set. Bengali is used to transcribe quite a few languages of which the most prominent are Assamese and Manipuri. The former differs from Bengali in a few consonant characters. The same is the case with Manipuri which today is also written in Meetei Mayek.

Principal Neo-Brāhmī Languages South Asian Scripts.

- *Oriya* [*Odia*] can be traced back to the Ashokan inscriptions: 3rd century B.C.E. Because of the prevalence of a large number of tribal languages belonging to the Munda and Dravidian families in the state of Odisha (Orissa), the Oriya script is used in writing these languages.
- *Sinhala* used for writing Sinhala language and at times also Pali, is derived from Brāhmī as early as the third-second century B.C.E. Although it belongs historically to the Northern family, it has been considerably influenced by the early Grantha script of South India.

- *Tamil (also spelt as "Tamizh")* More than any other script derived from Brāhmī, it is highly alphabetical in nature and admits no ligatures with the exception of two consonant conjuncts. Apart from being the official language of Tamil Nadu, Tamil is also an official and national language of Sri Lanka and one of the official languages of Singapore. Tamil is also spoken by significant minorities in Malaysia, England, Mauritius, Canada, South Africa, Fiji, Indonesia, as well as emigrant communities around the world.

- *Kannada* and *Telugu* are closely related scripts used to write two Dravidian languages: Kannada in the state of Karnataka, and Telugu in Telangana and Andhra. Over the centuries, Brāhmī evolved with marked characteristics in the south. Around the tenth century, these crystallised into the Old Kannada script, used where both Kannada and Telugu are now spoken. By around 1500, this script divided into Kannada and Telugu. As a result, there are very few differences between these two scripts.

- *Malayalam*: Subject to reforms, modern Malayalam has introduced alphabetic writing into the script, although the main structure of Malayalam still adheres to the akṣara.

Target Scripts
Justification of the choice

From this vast plethora of Brāhmī based languages, the Neo- Brāhmī Group has restricted itself to the scripts and languages in the slide that follows.

The choice has been determined by two factors.

- The complexity of the task. Unlike any other group mandated by ICANN, this group has to work not only with a large number of scripts but major languages used by these scripts. Thus Devanāgarī itself covers a large number of languages: 11 official not to mention other languages of India as well as outside India.
- Once the target scripts have been identified, languages using these scripts have to be delimited. ICANN has used the yard-stick of EGIDS (Expanded Graded Intergenerational Disruption Scale): a tool used to measure the status of a language in terms of endangerment or development. Only languages with an EGIDS scale of 1-4 will be admitted.
- *EGIDS is an expansion of Joshua Fishman's Graded Intergenerational Disruption Scale. When GIDS was applied to languages outside of Europe, problems arose. Paul Lewis and Gary Simons took the task of expanding GIDS so that any language in the world could be scored using it, hence the Expanded Intergenerational Disruption Scale. To do this they combined GIDS with UNESCO's six level scale (Brenzinger et. al. 2003). By combining the two scales, they came up with a 13 level scale that is rich enough in every category to score any language in the world. Each level has a number and label, given in the table in the next slide.* (Excerpted from Wiki: http://surveywiki.info/index.php/EGIDS)

Target Scripts Justification of the choice: EGIDS SCALE

Level	Label	Explanation
0	International	The language is widely used between nations in trade, knowledge exchange, and international policy.
1	National	The language is used in education, work, mass media, and government at the national level.
2	Provincial	The language is used in education, work, mass media, and government within major administrative subdivisions of a nation.
3	Wider Communication	The language is used in work and mass media without official status to transcend language differences across a region.
4	Educational	The language is in vigorous use, with standardization and literature being sustained through a widespread system of institutionally supported education.
5	Developing	The language is in vigorous use, with literature in a standardized form being used by some though this is not yet widespread or sustainable.
6a	Vigorous	The language is used for face-to-face communication by all generations and the situation is sustainable.
6b	Threatened	The language is used for face-to-face communication within all generations, but it is losing users.
7	Shifting	The child-bearing generation can use the language among themselves, but it is not being transmitted to children.
8a	Moribund	The only remaining active users of the language are members of the grandparent generation and older.
8b	Nearly Extinct	The only remaining users of the language are members of the grandparent generation or older who have little opportunity to use the language.
9	Dormant	The language serves as a reminder of heritage identity for an ethnic community, but no one has more than symbolic proficiency.
10	Extinct	The language is no longer used and no one retains a sense of ethnic identity associated with the language.

© SIL

Target Scripts: Justification of the choice

The following 21 languages have been targeted as part of the Neo- Brāhmī initiative. Since the group is community driven, languages can be added to by the community, provided approval of ICANN is obtained and the language in question complies with the EGIDs criteria.

Script	Language
Bengali	Assamese – অসমীয়া
Bengali	Bengali – বাংলা
Bengali	Manipuri – মনিপুরি
Devanāgarī	Bodo – बड़ो
Devanāgarī	Dogri – डोगरी
Devanāgarī	Hindi – हिन्दी
Devanāgarī	Kashmiri – कॉशुर, Kāšur, Koshur
Devanāgarī	Konkani – कोंकणी,
Devanāgarī	Maithili – मैथिली, মৈথিলী
Devanāgarī	Marathi – मराठी
Devanāgarī	Nepali – नेपाली
Devanāgarī	Sanskrit – संस्कृतम्, संस्कृतावाक्
Devanāgarī	Santali/Santhali – संथाली
Devanāgarī	Sindhi – सिंधी
Gujarati	Gujarati – ગુજરાતી
Gurumukhi	Punjabi – ਪੰਜਾਬੀ
Kannada	Kannada – ಕನ್ನಡ
Malayalam	Malayalam – മലയാളം
Oriya(Odia)	Odia – ଓଡ଼ିଆ
Tamil	Tamil – தமிழ், Tamizh
Telugu	Telugu – తెలుగు

Neo-Brāhmī LGR
Pan-Asian and not Pan-Indian

- Due to societal and commercial factors, quite a few languages which constitute the target languages of the group are also used as official languages elsewhere.
- Nepali and Maithili are official languages of India and also of the Democratic republic of Nepal. The same is the case with Bengali which is shared with Bangladesh. Hindi is the official language of Fiji and is widely used in Mauritius. Tamil is shared with Singapore and Malaysia.
- Some of these countries have made separate proposals to ICANN as is the case of Singapore and Bangladesh.
- Eventually a dialogue with these countries sharing a common language and those with a script sub-stratum such as Sinhala needs to be opened up.

धन्यवाद !