



अकारं सुखः सर्वव्यापकम्

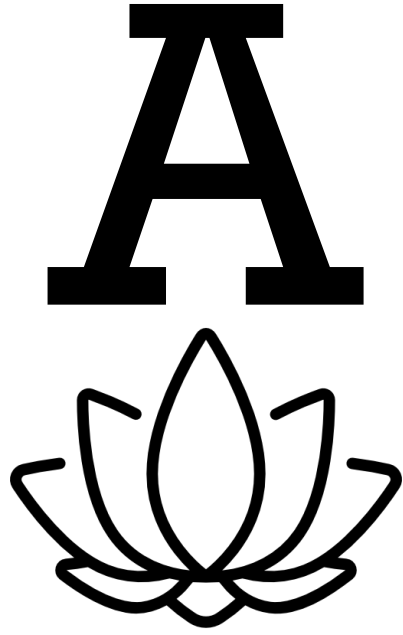
UNVEILING SCRIPTS:

Transliteration with
Aksharamukha

Vinodh Rajan

vinodh@virtualvinodh.com

<http://www.virtualvinodh.com>



Akāro Mukhaḥ Sarvadharmāṇaṃ

UNVEILING SCRIPTS:

Transliteration with
Aksharamukha

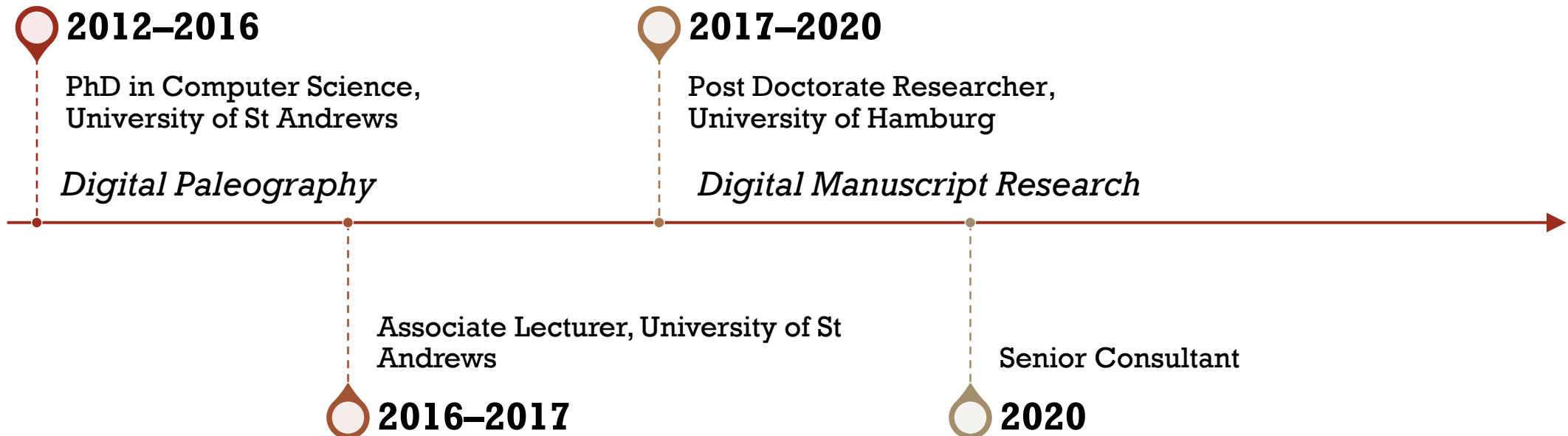
Vinodh Rajan

vinodh@virtualvinodh.com

<http://www.virtualvinodh.com>

(The letter) A is the door to all phenomena

ABOUT ME



TRANSLITERATION/ROMANIZATION

- Translation: Conversion between languages
 - मम नाम विनोद अस्ति → My name is Vinodh
- Transliteration: Conversion between scripts
 - मम नाम विनोद अस्ति → Mama nāma vinoda
- Romanization is type of transliteration where a script is rendered in Roman characters (very often with diacritics)
 - ISO 15919 for Indic, ISO 233 Arabic, ISO 259 Hebrew
 - Library of Congress Romanization

AKSHARAMUKHA

- Transliteration tool that supports conversion between 121 scripts
 - Indic
 - Semitic
 - Kana
- Supports 21 different romanizations
- Support reversible transliteration among major Indian scripts

akṣaramukha

अक्षरमुख

ಅಕ್ಷರಮುಖ

ਅਕਸ਼ਰਮੁਖ

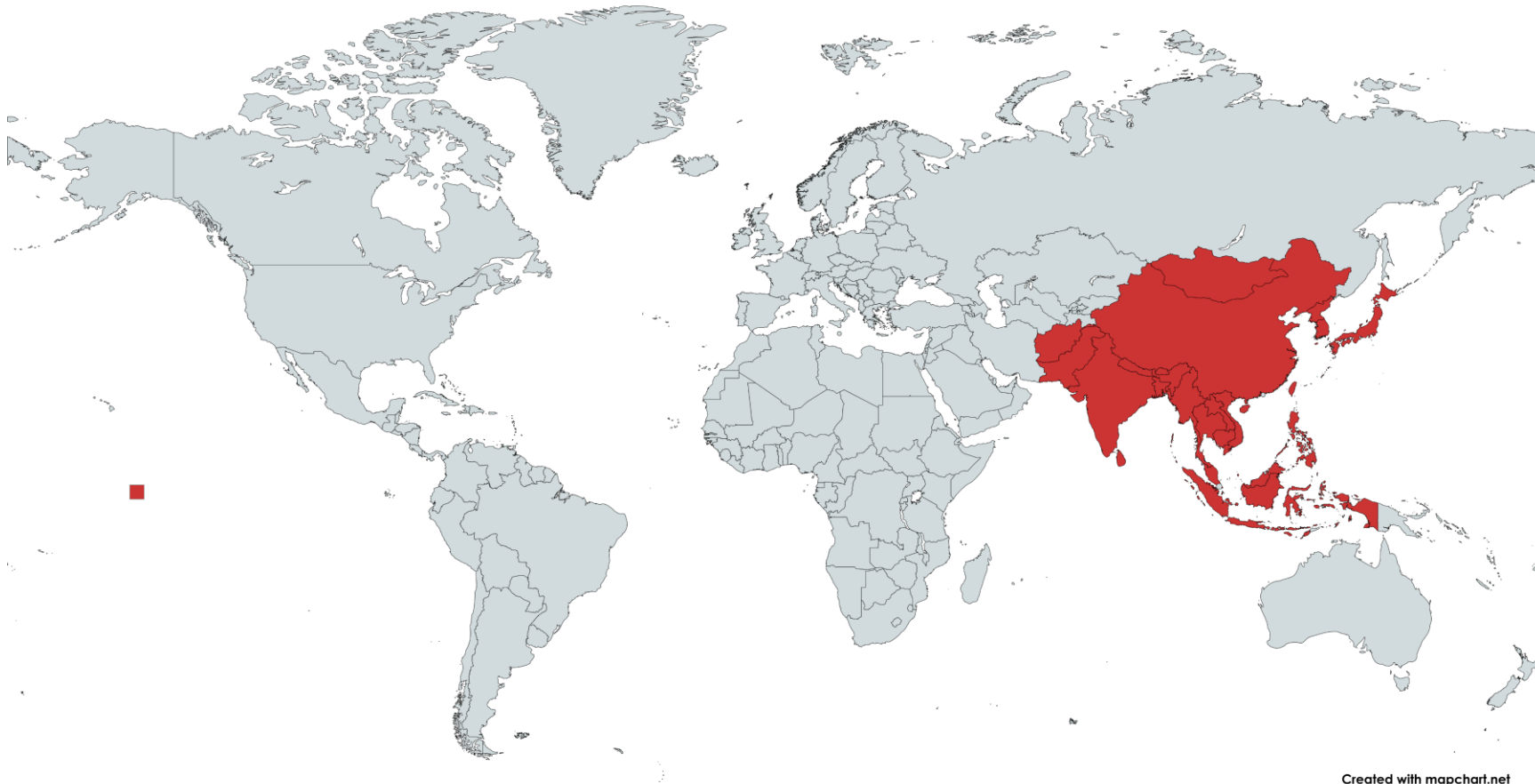
אַכְּשָׁרְמוּכָה

أَكْشَرْمُكْ

ਅਕਸ਼ਰਮੁਖ

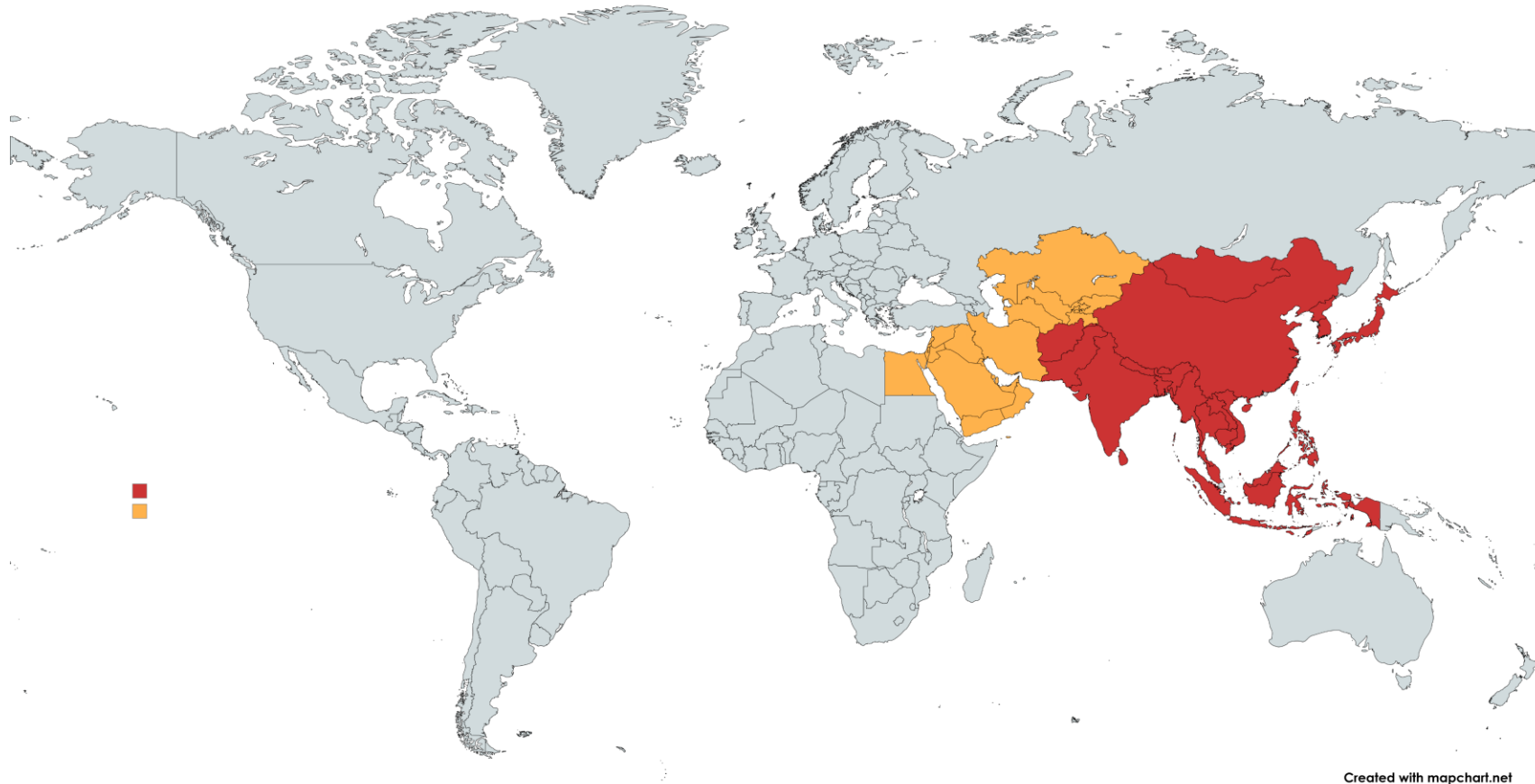
ਅਕਸ਼ਰਮੁਖ

INITIAL SCOPE



Created with mapchart.net

EXPANDED SCOPE



Created with mapchart.net

MOTIVATION

“ [...] *mantras* transcribed in Chinese are but a semblance of the original Sanskrit, how could such transcription ever be compared to the original text? [...] ”

Chinese monk I-ching, 7th century CE

- Enabling liturgical texts (mostly Sanskrit/Pali) to be accessible to everybody
- Enabling texts to be accessible to anyone from a Japanese Shingon Buddhist monk (using Siddham) to a person in Bali (using Balinese) to a person in Cambodia (using Khmer)
- Preserving the diversity of scripts without resorting to a “standardized” script
- Making “languages” transparent to scripts

TRANSLITERATION

సిద్ధార్థుడు కపిలవస్తు దేశానికీ చెందిన లుంబిని పట్టణంలో జన్మించాడు.

siddhārdhuḍu kapilavastu dēśānikī cendina lumbini paṭṭaṇaṃlō janmiñcāḍu.

సిద్ధార్థుడు కపిలవస్తు దేశానికీ చెందిన లుంబిని పట్టణంలో జన్మించాడు.

सिद्धार्थ कपिलवस्तु देशाనికి चेंदिन लुम्बिनी पट्टणंल जन्मिंचाडु.

ਸਿੱਧਾਰਥੁ ਕਪਿਲਵਸਤੁ ਦੇਸ਼ਾਨਿਕੀ ਚੇਂਦਿਨ ਲੁਮ੍ਬਿਨੀ ਪਟ੍ਟਣਾਨਲੋ ਜਨਮਿੰਚਾਡੁ.

سیدھا رھڈ کپلو ست دیشانکی چیندین لنبین پٹنل جمنچاڈ

TRANSLITERATION

ויאמר אלהים אל משה אהיה אשר אהיה ויאמר כה תאמר לבני ישראל אהיה שלחני אליכם

wy'mr 'lhym 'l mšh 'hyh 'sr 'hyh wy'mr kh t'mr lbny ysr'l 'hyh šlhny 'lykm

וַיֹּאמֶר אֱלֹהִים אֶל מֹשֶׁה אֶהְיֶה אֲשֶׁר אֶהְיֶה וַיֹּאמֶר כֹּה תֹאמַר לְבְנֵי יִשְׂרָאֵל אֶהְיֶה שְׁלַחֲנִי אֵלֵיכֶם

וַיֹּאמֶר אֱלֹהִים אֶל מֹשֶׁה אֶהְיֶה אֲשֶׁר אֶהְיֶה וַיֹּאמֶר כֹּה תֹאמַר לְבְנֵי יִשְׂרָאֵל אֶהְיֶה שְׁלַחֲנִי אֵלֵיכֶם

וַיֹּאמֶר אֱלֹהִים אֶל מֹשֶׁה אֶהְיֶה אֲשֶׁר אֶהְיֶה וַיֹּאמֶר כֹּה תֹאמַר לְבְנֵי יִשְׂרָאֵל אֶהְיֶה שְׁלַחֲנִי אֵלֵיכֶם

TRANSLITERATION

וַיֹּאמֶר אֱלֹהִים אֶל מֹשֶׁה אֲהִי אֲשֶׁר אֲהִי וַיֹּאמֶר כֹּה תֹאמַר לִבְנֵי יִשְׂרָאֵל אֲהִי אֲשֶׁר אֲהִיכֶם

way:o'mer ělohīm 'el moṣ̥eh 'ehyeh aṣ̣er 'ehyeh way:o'mer koh to'mar livnēy yiṣ̣rā'el
'ehyeh ṣ̣lāḥanī ālēykem

వయ్యోమేర్ ఔలోహీమ్ ఔల్ మోషే ఔహొ అషేర్ ఔహొ వయ్యోమేర్ కో తోమర్ లిన్ఱె యిస్ర్రాఁల్ ఔహొ షేలాఖనీ అలేఖేమ్

వయ్యోమేర్ ఎలోహీం ఎల్ మొషే ఎహ్యో అషేర్ ఎహ్యో వయ్యోమేర్ కొ తొమర్ లివ్నె యిస్ర్రాఁల్
ఎహ్యో షేలాఖనీ అలేఖేం

(vayyōmēr ělohīm ěl mōṣ̥e ěhyě aṣ̣er ěhyě vayyōmēr kō tōmar livne yisrāel ěhyě
ṣ̣ēlākhanī alekhēm)

MOTIVATION

- Create a tool that does transliteration between two scripts
 - Simple
 - Usable
 - Customizable
 - Supports minority scripts
 - Not need for special fonts to be installed
 - Ability to download the output as an image

FEATURES

- By default adheres to native orthographic conventions
- Provides extensive customization options
- Supports multiple input format : Text, File, Images
- Ability to work with multiple scripts
- Contains script information and (customizable) comparison tables
- Extensions: Python and Web Plugins
- Provides Input methods

CHALLENGES

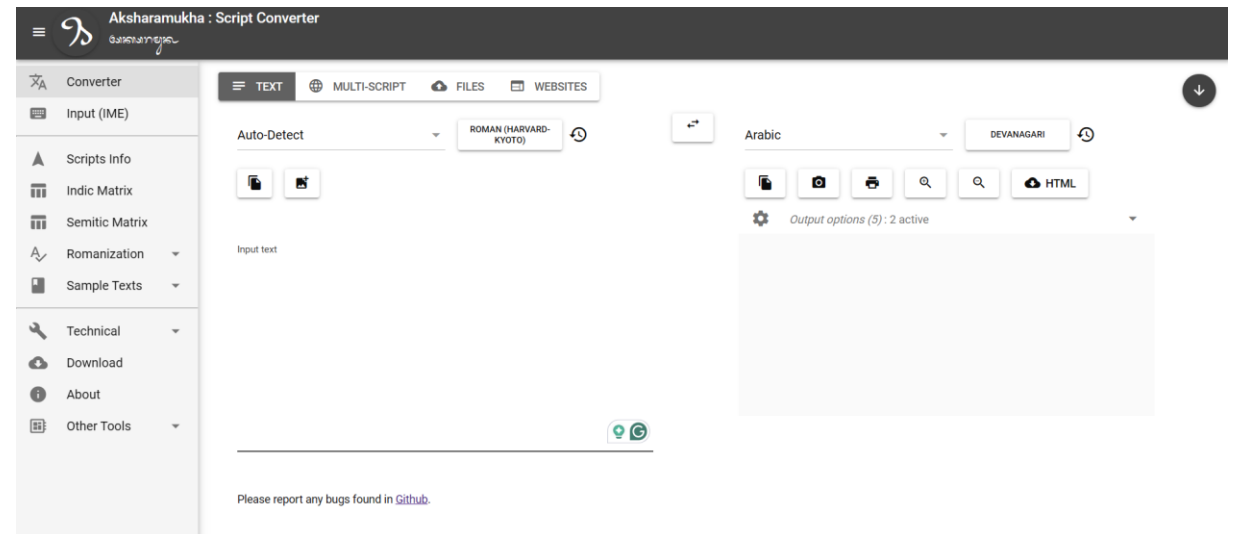
- Transliteration is not always straight forward when you've 121 pairs to handle
- Handling missing characters
- Facilitating lossless round-trip conversion
- Implementing orthographic conventions : Gemination etc.
- Deciding between Phonetic vs Etymological mapping
 - Arabic ج is analogous to Hebrew ך but phonetically its is equivalent to ڭ
 - Hebrew ך → ج ڭ or even ق ?

LIBRARY OF CONGRESS ROMANIZATION

- Aksharamukha supports LoC romanization for about 21 scripts out of 42 scripts included in Aksharamukha that have an LoC scheme
- It's not a lot but it's a good start. Full list here: <https://www.aksharamukha.com/loc>
- Supports MARC-8 output (decomposed Unicode diacritics)
- Support language specific needs such as syllable segmentation for scripts that require them e.g. *Burmese, Khmer*




<https://www.aksharamukha.com>



FUTURE WORK

- Some LoC romanization schemes are not reversible
 - संघ सङ्घ → saṅgha saṅgha
- Add ability to reverse by guessing based on orthographical conventions
 - Saṅgha → सङ्घ (if Sanskrit), संघ (if Hindi)
- Add more scripts!
- Resolve bugs. No software is perfect.
- If you find something that doesn't work as expected, please report it



*viśvāmitramācāryamevamāha - katamāṃ me
bho upādhyāya lipiṃ śikṣāpayasi?
brāhmīkharoṣṭīpuṣkarasāriṃ aṅgalipiṃ
vaṅgalipiṃ [...] cīnalipiṃ lūnalipiṃ hūṇalipiṃ
[...] āsāṃ bho upādhyāya catuṣṣaṣṭīlipīnāṃ
katamāṃ tvam śiṣyāpayiṣyasi?*

[He then] addressed [his] teacher Vishvamitra. Which script, O master, will you be teaching me? [The] Brahmi [script], [the] Kharoshthi [script], [the] Pushkarasari [script], the Anga script, the Vanga script, [...] the Chinese script, the Luna script, the Hun script [...] Master, which of those 64 scripts will you teach me?

— The young Buddha to his teacher (Lalitavistara Sutra)