ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
Please fill all the sections A, B and C below.
(Please read Principles and Procedures Document for guidelines and details before filling this form.)
See http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html for latest *Form.*
See http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html for latest *Principles and Procedures* document.
See http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html for latest roadmaps.

A. Administrative

1. **Title: Proposal for the Encoding of Brāhmī in Plane 1 of ISO/IEC 10646**
2. Requesters' names: Stefan Baums, Andrew Glass.
3. Requester type (Member body/Liaison/Individual contribution): Individual contribution.
4. Submission date: DRAFT 27 July 2003.
5. Requester's reference (if applicable): N/A.
6. This is a complete proposal.

B. Technical - General

1. This proposal is for a new script (set of characters).
   Proposed name of script: Brāhmī.
2. Number of characters in proposal: 120.
3. Proposed category (see section II, Character Categories): C.
4. Proposed Level of Implementation (1, 2 or 3) (see clause 14, ISO/IEC 10646-1: 2000): 3.
   Is a rationale provided for the choice? Yes.
   If Yes, reference: Combining marks are used.
5. 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 ISO/IEC 10646-1: 2000? Yes.
   b. Are the character shapes attached in a legible form suitable for review? Yes.
6. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard? Stefan Baums.
   If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used: Department of Asian Languages and Literature, University of Washington, Box 353521, Seattle, WA 98195-3521, USA, baums@u.washington.edu. GIMP, PfaEdit.
7. 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.
8. 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.
9. 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 http://www.unicode.org/Public/UNIDATA/UnicodeCharacterDatabase.html 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? No.
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.)? Yes.
   If YES, with whom? Richard Salomon, Lore Sander, Jost Gippert.
   If YES, available relevant documents: .
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Indologists, Iranianists, Turkologists, Indo-Europeanists, South-East-Asianist.
4. The context of use for the proposed characters (type of use; common or rare): Scholarly; common.
5. Are the proposed characters in current use by the user community? Yes.

[1] Form number: N2352-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09)

If Yes, where? Reference: Palaeographic, epigraphic and philological studies; text editions.

6. After giving due considerations to the principles in *Principles and Procedures document* (a WG 2 standing document) must the proposed characters be entirely in the BMP? No.

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? Yes.

8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? No.

9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? No.

10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? Yes.

    If Yes, is a rationale for its inclusion provided? Yes.

        If Yes, reference: See below.

11. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)? Yes.

    If Yes, is a rationale for such use provided? Yes.

        If Yes, reference: See below.

    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? Yes.

        If Yes, reference: See below.

12. Does the proposal contain characters with any special properties such as control function or similar semantics? Yes.

    If Yes, describe in detail (include attachment if necessary). U+1104D BRAHMI SIGN VIRAMA, see below.

13. Does the proposal contain any Ideographic compatibility character(s)? No.

Proposal for the Encoding of Brāhmī in Plane 1 of ISO/IEC 10646

1. Brāhmī Unification

In spite of superficial historical and regional variation in the form of letters and their combinations, the members of the pre-modern Brāhmī script family agree very closely in character repertoire and systemic principles. The variation that does exist is of a gradual nature that would make it a very difficult and rather arbitrary task to break the Brāhmī script continuum down into subvarieties. While in the study of Brāhmī palaeography, questions of subclassification and variation do need to be discussed, we are convinced that in digital form this variation is most suitably represented at the font level, not at the encoding level. Setting one particular conception of Brāhmī subvariation in stone by making it the basis of a non-unified character coding would hinder rather than help palaeographical study as subclassification schemes continue to be revised. It must also be kept in mind that in pre-modern India there was to a very large extent no natural connection between script varieties on the one hand and languages and their texts on the other: any given script variety would typically be used for the writing down of texts in multiple languages (such as Sanskrit and one or more regional languages), and any given text would be written in different parts of India in the respective regional scripts. Therefore the Indological community – the main potential users of a Brāhmī character coding – typically has to handle manuscript and epigraphical material in a multitude of script varieties in their investigation of a single text or group of texts. An artificially non-unified encoding of the written source material for this sort of study would greatly complicate searching and general data-processing.

2. Overview of the History of pre-modern Brāhmī

The earliest examples of writing from historical India are the edicts of emperor Aśoka from the third c. BCE. Most of his inscriptions are in the Brāhmī writing system, but in the Indian northwest Kharoṣṭhī, Aramaic and Greek are used as well. It would appear that the earliest known form of Brāhmī presupposes the existence of Kharoṣṭhī: Brāhmī follows the same system of vowel marking as Kharoṣṭhī, but has a greater number of distinct vowel signs that allow for a much better representation of Indian speech; and Kharoṣṭhī has clear historical associations (with the Aramaic script) that Brāhmī lacks. It has been suggested that the Brāhmī script was specially invented for use in the royal inscriptions of Aśoka or documents of their kind, on the basis of an acquaintance with the Kharoṣṭhī and maybe also the Aramaic or Greek scripts. The name 'Brāhmī' has been applied to this script family by modern scholars and is taken from the list of scripts that the young Buddha is claimed to have mastered in the *Lalitavistara*; the first script on this long list is called *brāhmī* and said to be written from left to right, while the second is called *kharoṣṭhī* and said to be written from right to left.

The further historical development of the Brāhmī script is characterised by very gradual changes in the forms of letters conditioned by cursivisation and modification of stroke order, and by changes in the writing utensils used. The characteristic headmarks of the modern Devanāgarī and Bengali scripts, for instance, have their origin in the mark left where a reed pen first touches the writing surface, and the trend towards round letter forms in the southern varieties of Brāhmī is attributed to the southern technique of incising letters into palm leaves, where straight lines would have tended to split the leaf.

One widely used system of palaeographical subclassification is that developed by A. H. Dani, distinguishing Old, Middle, and Late Brāhmī periods, Transitional Scripts, and the modern Indian scripts. While spreading towards southern India in the Old Brāhmī period (third to first c. BCE), the script was subject to experimental and rather shortlived systemic innovations attested in the Old Tamil and Bhattiprolu inscriptions (see below). In the Middle Brāhmī period (first to third c. CE), regional variation increases; Dani distinguishes between Mathurā, Kauśāmbī, Western Deccan and Eastern Deccan styles. Brāhmī is for the first time being used to represent Sanskrit, and for this purpose four new characters are added to the script (ҳ / ҫ r̥, ҙ / ҭ au, ҩ ḥ, and ҽ ṅa). A special device is introduced for the marking of vowelless consonants, used both for Sanskrit, where it is called *virāma* and first occurs in manuscripts of the first c. CE, and for Tamil, where it is called *puḷḷi* and attested in inscriptions from the second c. CE (Mahadevan 2003, p. 198). In the course of trade relations and cultural exchange, the Brāhmī script is being exported to Central Asia and Southeast Asia. For several centuries, Indian forms of the script continue being used in both these regions, primarily for the writing of Sanskrit texts. It is first during the Late Brāhmī period (fourth to seventh c. CE) that distinct Central Asian and Southeast Asian forms of Brāhmī develop, which then also begin being used for the writing of local languages. While the Central Asian tradition of Brāhmī comes to an end with the Muslim invasions of the region at the end of the first millenium, the Southeast Asian forms of Brāhmī develop further into the modern Southeast Asian scripts. In the period of the Transitional Scripts (seventh to tenth c. CE), the Indian Northwest sees the emergence of the proto-Śāradā form of Brāhmī that became the precursor of Śāradā and other regional scripts such as Takri and Landa, which inspired the development of the modern Gurmukhi script. In the rest of northern India, a style called Siddhamātr̥kā predominated that gave rise to the modern Devanāgarī and Bengali scripts. In the Deccan, a proto-Kannada-Telugu script began to take form, while further south the Grantha script developed for the writing of Sanskrit, and the Vaṭṭeḻuttu and Tamil scripts for the writing of Tamil.

3. General properties of the Brāhmī script

The Brāhmī script shares many properties with Devanāgarī and its other descendants. Lines are usually written from left to right and pages filled from top to bottom. In almost all varieties of Brāhmī (but see below on Tamil and Bhattiprolu Brāhmī), the basic consonant graphemes denote the consonant in combination with an inherent *a* vowel. The presence of other vowels is indicated by adding vowel diacritics to the base consonant, as illustrated below from the Gilgit-Bamiyan type I variety of Brāhmī (sixth/seventh c. CE, Northwestern India):

ta	tā	ti	tī	tu	tū	tr̥	te	tai	to	tau
U+1101F	U+1101F, U+11035	U+1101F, U+11036	U+1101F, U+11037	U+1101F, U+11038	U+1101F, U+11039	U+1101F, U+1103A	U+1101F, U+1103E	U+1101F, U+1103F	U+1101F, U+11040	U+1101F, U+11041

A sequence of consonants without intervening vowels is written as a consonant ligature. As with the other Indic scripts, these consonant ligatures are to be encoded with the help of U+1104D BRAHMI SIGN VIRAMA. It is to be noted that up to a very late date, Brāhmī used vertical conjuncts exclusively; there is thus no parallel series of 'half-consonants' as in Devanāgarī and other modern scripts. Consonant ligatures are written from top-left to bottom-right:

𑀢𑁆𑀫	𑀢𑁆𑀲	𑀢𑁆𑀓𑁆𑀱	𑀤𑁆𑀕𑁆𑀭	𑀰𑁆𑀫	𑀲𑁆𑀣𑀸
tma	*tsa*	*tkṣa*	*dgr*	*śma*	*sthā*
		U+1101F,			U+1102F,
U+1101F,	U+1101F,	U+1104D,	U+11021,	U+1102D,	U+1104D,
U+1104D,	U+1104D,	U+11010,	U+1104D,	U+1104D,	U+11020,
U+11028	U+1102F	U+1104D,	U+11012,	U+11028	U+11035
		U+1102E	U+1103A		

Pre- and postconsonantal *r* and postconsonantal *y* assume special reduced shapes in all but the earliest varieties of Brāhmī; the *kṣa* and *jña* ligatures, however, are often transparent:

𑀭𑁆𑀢𑀼	𑀢𑁆𑀭	𑀢𑁆𑀬	𑀭𑁆𑀬	𑀓𑁆𑀱	𑀚𑁆𑀜
rtu	*tra*	*tya*	*rya*	*kṣa*	*jña*
U+1102A,					
U+1104D,	U+1101F,	U+1101F,	U+1102A,	U+11010,	U+11017,
U+1101F,	U+1104D,	U+1104D,	U+1104D,	U+1104D,	U+1104D,
U+11038	U+1102A	U+11029	U+11029	U+1102E	U+11019

When a consonant without inherent vowel cannot be written as non-final part of a ligature, such as when that consonant occurs at the end of a verse or paragraph, a visible *virāma* device is used. This device consists primarily of writing the vowelless consonant smaller and lower than other consonants, and often also of drawing a connecting line from the vowelless consonant to the preceding *akṣara*. Secondarily, a short horizontal line is frequently added above the vowelless consonant; it is this horizontal line that developed into the visible *virāma* marks of the modern Brāhmī-derived scripts.

𑀘𑀢𑁆	𑀚𑁂𑀢𑁆
cat	*jet*
	U+11017,
U+11015,	U+1103E,
U+1101F,	U+1101F,
U+1104D	U+1104D

The *anusvāra* sign (U+11045) is used to indicate that a vowel is nasalised (when the next syllable starts with a fricative), or that it is followed by a nasal segment (when the next syllable starts with a stop). The need for a separate encoding of *candrabindu* (indicating only nasalisation of a vowel) could not yet be demonstrated, but the codepoint following *anusvāra* has been left unassigned in case the need should arise. The *visarga* sign (U+11047) is used to write syllable-final voiceless [h]. The velar and labial allophones of [h], followed by voiceless velar and labial stops respectively, are sometimes written with the separate signs *jihvāmūlīya* and *upadhmānīya* (U+11048 and U+11049); in contrast to *visarga*, these two signs are not combining diacritics, but behave like ordinary consonant signs, entering into ligatures with the following stop. (The third and fourth illustrations in the following table are from a Gupta dynasty manuscript of the fourth/fifth c. CE).

𑀢𑀁	𑀢𑀂	𑀳𑁆𑀓	𑀳𑁆𑀨
taṃ	*taḥ*	*ḫka*	*ḫpha*
		U+11048,	U+11049,
U+1101F,	U+1101F,	U+1104D,	U+1104D,
U+11045	U+11047	U+11010	U+11025

It should be noted that the independent vowel signs *ṝ*, *ḷ* and *ḹ* and the dependent vowel signs *ḷ* and *ḹ* hardly ever occur in ordinary written texts, and therefore could not be illustrated in the code charts. They are, however, recognised by the indigenous Indian systems of grammar,

and therefore always have the potential of being written. At least dependent *ḷ* and *ḹ* occur in manuscript abecedaries, and their glyphs will eventually be added to the code charts. The situation is exactly parallel to that of the corresponding characters in Devanāgarī, with the exception that the Devanāgarī characters had already been sanctioned by the ISCII standard and therefore had to be included in Unicode, and that illustrative glyphs were more readily available.

The Brāhmī script has separate number signs not only for the digits from 1 to 9, but also for the tens from 10 to 90, and for 100, 200, 300, 1000, 2000 and 3000 (no attestation has yet been found of the last, and therefore no codepoint has yet been assigned to it). Numbers are written additively, with higher number signs preceding lower ones. Multiples of 100 higher than 300, and of 1000 higher than 3000, are expressed multiplicatively, with the multiplier following and forming a ligature with *100* or *1000*; we suggest that these ligatures be encoded with with ᴢᴇʀᴏ ᴡɪᴅᴛʜ ᴊᴏɪɴᴇʀ (U+200D). (The first five illustrations in the following table are based on a Gupta dynasty manuscript; the last two are from Western Kṣatrapa coin legends of the first to fourth c. CE.)

∝𝒫	𝒢-	𝓃	𝓃	𝓃𝔂	𝟫	𝔀𝔀
10 6 (= 16)	*50 1* (= 51)	*100*	*200*	*100 4* (= 104)	*1000*	*1000-4* (= 4000)
U+1105A, U+11056	U+1105E, U+11051	U+11063	U+11064	U+11063, U+11054	U+11064	U+11064, U+200D, U+22054

Later in the history of Brāhmī, a special sign for zero (U+11050) was invented, and the positional system came gradually into use.

Seven punctuation marks should be encoded, namely single (ˌ, U+11070) and double (ˌˌ, U+11071) *daṇḍa*, delimiting clauses and verses; dot (ˌ, U+11072), double dot (ˌ, U+11073) and horizontal line (ˌ, U+11074), delimiting smaller textual units; and the crescent (ᴈ, U+11075) and lotus (❀, U+11076) marks, delimiting larger textual units. The scribes of Brāhmī manuscripts use additional devices, such as horizontal wavy lines and larger floral designs, to structure their texts, but these are of very disparate appearance and often their shape and presence is determined by physical features of the manuscript. Therefore they should be considered graphical elements rather than punctuation proper, comparable to vignettes in European manuscripts and prints.

4. Tamil Brāhmī

In the second c. BCE, as Brāhmī spread southwards, speakers of Old Tamil became acquainted with it and adapted it to the writing of their own language. The Tamil form of Brāhmī is known to us from a number of inscriptions donating caves to Jaina monastic communities, mostly in southern Tamil Nadu; from pottery graffiti found at Arikamedu, Kodumanal and other ancient trading sites; and from coin legends and inscriptions on objects such as seals and rings. In contrast to the Middle Indo-Aryan dialects for which Brāhmī had been originally invented and used so far, the Tamil language has word-final consonants that needed to be represented in the writing system. In its first phase of development (Early Tamil Brāhmī, second c. BCE – first c. CE), two competing modifications of Brāhmī orthography were used to achieve this aim. The one system (Mahadevan 2003's 'TB-I') does away with the inherent vowel of Brāhmī consonant signs, using the vowel *mātrā ā* to represent both short and long [a] / [aː]; consonant signs without *mātrā* always represent the

bare consonant in this orthography. In the second orthographic system (Mahadevan's 'TB-II'), the *ā mātrā* always represents long [aː], whereas vowelless consonant signs can be read either with inherent short [a] or as bare consonants, depending on the context. The element of ambiguity in both these systems (of *ā* in TB-I and of bare consonant signs in TB-II), as well as pressure to conform with regular forms of Brāhmī that had been adopted in neighbouring regions, led to a further orthographic modification (Late Tamil Brāhmī, second – fourth c. CE, Mahadevan's 'TB-III') with the adoption of the *puḷḷi* diacritic to unambiguously mark vowelless consonants. *Puḷḷi* takes the form of a dot above or in the upper part of the *akṣara*. In addition to this normal *virāma* function, *puḷḷi* is also used with the vowels *e* and *o* in order to mark them as short: in contrast to Sanskrit and most Middle-Indo-Aryan dialects, the Dravidian languages have short as well as long *e* and *o* phonemes. Just as in other forms of Brāhmī, short [a] is always inherent in TB-III consonant signs, and *ā* always means long [aː].

The orthographic peculiarities of Old Tamil Brāhmī do not concern the elements of the writing system itself, but are a matter of the conventional phonetic interpretation of these elements. The encoding of Old Tamil Brāhmī should not reflect this phonetic interpretation, but should be based on what is actually written; bare *akṣaras* and *akṣaras* with *ā mātrā* should be encoded as such, just as in other varieties of Brāhmī. This is in accordance with Mahadevan 2003, who in his edition of the Old Tamil inscriptions provides first a close transliteration (corresponding to the proposed computer encoding of Old Tamil Brāhmī) and then a phonetic transcription (the following example is the second line of inscription no. 1, on p. 315, illustrating the TB-I system):

ku va a ṉa ke dha ma mā ma
kuv aṉkē dhammam

A similar encoding principle obtains already in the case of Devanāgarī as used for Hindi and of the Gurmukhi script, where by conventional phonetic interpretation morpheme-final bare *akṣaras* are pronounced vowelless without this being reflected at the encoding level. The two functions of Late Tamil Brāhmī *puḷḷi* can be subsumed under the heading of 'vowel reduction' (short to zero, and long to short), and *puḷḷi* should be encoded as U+1104D BRAHMI SIGN VIRAMA; the Brāhmī *virāma* character can thus follow both consonant characters and the vowel characters *e* and *o*, in contrast to the modern scripts' *virāma* characters (cf. below for another use in combination with vowels in the Uigur orthography).

For the representation of sounds particular to Dravidian, the makers of Old Tamil Brāhmī added four new consonant signs to the repertoire of Brāhmī: ℗ *ḷ*, ᴛʜ *ḻ*, ⸗ *ṟ* and ⸗ *ṉ*. The second of these, *ḻ*, is phonetically identical (a retroflex lateral) to the *ḷ* that somewhat later appears in north-Indian Brāhmī for the writing of Sanskrit, and that also occurs in the Bhattiprolu inscriptions. Moreover, both the Tamil Brāhmī and the Bhattiprolu *ḷ* are graphically derived from the regular letter *l*, the former by adding a hook to the lower right of *l*, the latter by mirroring *l* horizontally (while the north-Indian *ḷ* is derived from the letter *ḍ*). Old Tamil, Bhattiprolu and north-Indian *ḷ* should therefore all be encoded as U+11031. Additional codepoints are provided for *ḻ*, *ṟ* and *ṉ* in the positions U+11080 to U+11082.

5. Bhattiprolu Brāhmī

Ten short Middle Indo-Aryan inscriptions from the second c. BCE, found in a stūpa at

Bhattiprolu in Andhra Pradesh, show an orthography that seems to be derived from the Tamil Brāhmī system TB-I. To avoid the phonetic ambiguity of the latter's *ā mātrā* (standing for either [a] or [aː]), the Bhattiprolu inscriptions introduce a separate *mātrā* for long [aː] by adding a vertical stroke to the end of the *ā mātrā*: ᴄᴦ. Thus in these inscriptions, *ā* unambiguously means [a], and ᴄᴦ (here transliterated as *Ā*) means [aː]. (The following illustration is line 2 of inscription V in Bühler 1894; the reading follows Lüders 1912.)

hi rā ṇā kᴀ rā gᴀ mā ṇī pu to bū bo

hiraṇakāra   gāmaṇīputo   būbo

Puzzlingly, the main reason for abandoning inherent [a], namely the ability to write word-final consonants or non-homorganic consonant clusters conveniently, does not apply in the case of the Bhattiprolu inscriptions since Middle Indo-Aryan has neither of these phonetic features. This makes it likely that the dedicated long *Ā mātrā*, too, was first introduced in a Tamil context, and that the resulting system was only later imitated in Bhattiprolu. No such Tamil inscription has however been discovered yet.

The shapes of five Bhattiprolu letters (*gha*, *ja*, *ma*, *la* and *sa*) differ to a certain degree from those seen in other varieties of Old Brāhmī (the *ma*, for instance, is upside-down), but only in the case of *gha* (which is graphically derived from the unspirated *ga*) is there real innovation. Even *gha*, however, should be encoded as in other varieties of Brāhmī as its graphemic identity is not in doubt. The experimentation with letter shapes that we see in Bhattiprolu and other Old Brāhmī is entirely typical of early writing systems, such as the various Greek alphabets before the Athenian orthographic reform. The [ks] sound, for instance, was written X in the Western part of the Greek world and Ξ in Greece itself, a situation not unlike that of Bhattiprolu and regular *gha*.

6. Central Asian Brāhmī

It is not clear which Central Asian people first modified Brāhmī for the writing of their own language; both the Khotanese Saka (Hitch 1981) and the Tocharians (Sander 1986) have arguments in their favour. The Tocharians added a set of 11 new characters (the so-called *Fremdzeichen*, i.e., foreign or special signs) that differ from the corresponding regular Brāhmī characters by having inherent not an [a], but an [ə] sound, transliterated *ä*: ꙮ *ka*, ᴑ *ṭa*, ◇ *na*, ᴂ *pa*, ᴕ *ma*, ᴗ *ra*, ᴈ *la*, ᴝ *śa*, ᴔ *ṣa*, ᴖ *sa*, ᴕ *wa*. A diacritic double dot (ᴕ) indicates *ä* in combination with other consonants signs.

The Uigurs share the Tocharian special signs in their word-final use with *virāma*. They added six further signs to write special consonants of their own language: ꙮ *qa*, ᴕ *γa*, ᴕ *δa*, ᴕ *dza*, ᴕ *za*, and ᴕ *źa*. The Uigur short vowels *ä*, *ü* and *ö* are spelled *-ya-*, *-yu-* and *-yo-* postconsonantally. The long vowels *ǟ*, *ǖ* and *ȫ* are written like the short ones but with the addition of an *ā mātrā* (U+11035) to the same *akṣara*, which means that in the case of *ü* and *ö*, the *akṣara* carries not one but two vowel *mātrā*s (U+11038, U+11035 and U+11040, U+11035, respectively). The initial vowels *ä*, *ü*, *ö* and *ȫ* are written by adding *-ya-*, *-yu-*, *-ya-* and *-yo-* directly to the initial vowel signs *a* or *e*, *u*, *o* and *o*; this means that the resulting

complexes *aya-*, *eya-*, *uyu-*, *oya-* and *oyo-* are single *akṣara*s that should, on analogy with the postconsonantal vowels, be encoded with the control character U+1104D ʙʀᴀʜᴍɪ sɪɢɴ ᴠɪʀᴀᴍᴀ between the initial vowel character and the *-y-*:

𑀣/	𑀢	𑀤𑀺	𑀨	𑀨𑁀
aya	eya	uyu	oya	oyo
(= *ä*)	(= *ä*)	(= *ü*)	(= *ö*)	(= *ȫ*)
U+11000,	U+1100A,	U+11004,	U+1100C,	U+1100C,
U+1104D,	U+1104D,	U+1104D,	U+1104D,	U+1104D,
U+11029	U+11029	U+11029,	U+11029	U+11029,
		U+11038		U+11040

Uigur Brāhmī also employs the double dot diacritic ◌̈ to indicate high unrounded vowels.

The third language group in Central Asia that adopted the Brāhmī script for their own language are the Saka. Their Iranian language is attested in two varieties, Khotanese Saka and Tumshuqese Saka. The former dialect adds the diacritic double dot ◌̈ (U+11095) to the common Brāhmī repertoire, and shares with Uigur the un-Indian orthographic practice of adding two vowel *mātrā*s to a single *akṣara* for the writing of its set of falling diphthongs. Khotanese Saka also developed an alternative analytic way of writing word-initial vowels, using not the dedicated initial signs for all vowels, but just initial *a* as vowel bearer in combination with the various vowel *mātrā*s (cf. Hitch 1981, p. 42—44). The same system had been observed by Kharoṣṭhī, and was later reinvented (for some of their initial vowels) by the Gujarati and Devanāgarī scripts. In addition, Khotanese Saka makes use of a diacritic sign with the shape of a hook below the *akṣara* and of uncertain phonetic value; this sign has not yet been included in the proposed encoding pending further research.

Tumshuqese Saka employs a large number of special signs. The following manuscript sign list (published in Konow 1935) has twelve entries:
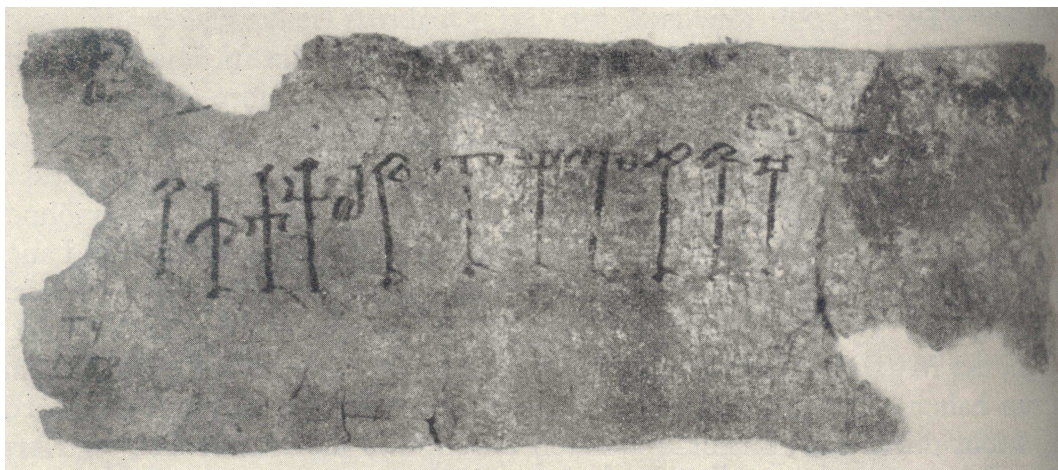
At least five of these signs (𐂀 *za*, 𐂁 *γa*, 𐂂 * źa*, 𐂃 *ḏa* and 𐂄 *dza*) are shared with Uigur and do not need to be encoded separately (their codepoints are U+1109D, U+1109A, U+1109E, U+1109B and U+1109C). Another three signs (nos. 3, 8 and 9 from the left) would appear to be mere copies of signs no. 2, 4 and 7 (𐂁 = 𐂁 *γa*, 𐂂 = 𐂂 *źa* and 𐂃 = 𐂃 *ḏa*), and are indeed not independently attested in Tumshuqese Saka manuscripts. The status of signs no. 5, 6 and 11 (𐂅, 𐂆 and 𐂇) has been disputed (Hitch 1981, pp. 67—77, analyses them as *la*, *khu* and *śu* instead of Konow's 1947 *źya*, *ẓa* and *gwa*), and they are therefore not included in the present

proposal. Sign no. 12 (Ⲧ), however, is generally agreed to be a genuine special character with the value *χśa*; it is included at codepoint U+110A2.

The Central Asian varieties of Brāhmī share a ligature *rra* that does not occur in Indian Brāhmī. Although *rra* tends to be treated as a unit in Khotanese Saka, probably representing a phoneme of that language distinct from the one written *ra*, it should be encoded as the ligature that, orthographically, it is.

7. Implementation and Usage

Is is anticipated that the main initial use of the Unicode Brāhmī encoding will be in the area of scholarly palaeographical work. Most of the fonts produced in this area of study will aim to reproduce a particular manuscript hand or epigraphic ductus as closely as possible. Every occurring *akṣara* instance (consonant-vowel-diacritic combination) will be assigned a single glyph in the font, and the use of combining vowel-sign glyphs and the like will be minimal. The main operation to be performed at the rendering level will therefore be the substitution of one particular *akṣara* glyph for a sequence of character code points, not the relative positioning of subparts of *akṣara*s as with modern Indic scripts. Most fonts produced for palaeographic purposes will not contain glyphs for every Brāhmī codepoint, and will usually not be applied to texts much different from the manuscripts or inscriptions based on which they were initially produced.

Ultimately, however, the production and distribution of comprehensive fall-back fonts for the main varieties of Brāhmī is desirable. These fonts will contain normalised glyph shapes, and in their case the use of combining glyphs for subparts of *akṣara*s is feasible. As with the other scripts included in the Unicode Standard, the memory representation of strings will follow their phonetic order. For most *akṣara*s in most varieties of Brāhmī, no display reordering such as for Devanāgarī *i* will be required, because the dependent vowel sign for *i* had not yet descended from its original position on top of the base consonant. Exceptions do, however, occur even in one and the same script, cf. Gilgit-Bamiyan type I ⟨ *dhi* with ⟨ *ti*; and in the medieval South Indian forms of Brāhmī, the *e* and *ai mātrā*s are regularly written on the left side of the *akṣara*.

It has been our aim to present a unified proposal for all pre-modern forms of Brāhmī, for the reasons set out at the beginning of this document. Looking back, possibly the strongest case for a separate encoding of a Brāhmī variety would have been Tamil Brāhmī due to the systemic characteristics that distinguish it from other forms of Brāhmī. As has been shown, however, the only way to encode the three subvarieties of Tamil Brāhmī (TB-I, TB-II and TB-III) uniformly and naturally is to regard the Tamil Brāhmī orthographic system as a matter of phonetic interpretation, not of character coding; any special encoding for this orthography would have separated TB-I and TB-II from TB-III, obscuring the historical development that after a period of experimentation reintegrates the Tamil variety into the mainstream of Brāhmī script history. The other varieties of Brāhmī diverge far less from the original model, and to unify their encoding should be even less controversial.

We strongly suggest that all historical documents written in a variety of Brāhmī be encoded following the codepoints and principles set out in this document. Additional characters that may become necessary for the encoding of future discoveries of Brāhmī texts can easily be added to the code range; no major additions are, however, expected.

It remains up to the user's discretion whether in individual cases his documents are most

naturally encoded using the Brāhmī code range or the code range of one of the modern Brāhmī-derived scripts, an issue similar to the linguistic dilemma of when exactly to start regarding texts as written in New Indo-Aryan instead of Middle Indo-Aryan. (It is worth pointing out again that this problem of decision would be exacerbated manifold if the historical varieties of Brāhmī were encoded in a non-unified manner.) In practice, the set of characters provided respectively by the Brāhmī range and by the modern-script ranges will have an influence on the user's decision. For example, an early Sri Lankan text containing the special Sinhalese vowel *ä* could not be encoded as Brāhmī, since the present proposal does not contain a codepoint for this vowel, but only as Sinhalese using the Unicode Sinhala code range (U+0D80 to U+0DFF). It is part of our responsibility to make this sort of delimitation imposed by the contents of the Brāhmī code range coincide as closely as possible with the boundaries suggested by linguistic and other scholarly criteria.

8. Sorting

Alphabetically ordered word lists (such as dictionaries) in the Brāhmī script are not preserved and maybe never existed. We do, however, know the traditional way of arranging the letters of the Brāhmī script from ancient abecedaries (*varṇamālā*s or *dvādaśākṣarī*s) which are based on phonetic principles. The sort order of the modern Indian scripts, as well as of Indologist transliteration, is based on the *varṇamālā* order, but varies in some details. The conjuncts *kṣ* and *jñ*, for instance, are considered so basic that they are included in their own right at the end of the ancient *varṇamālā*s; this is not imitated in modern usage.

It is most practical to specify the Brāhmī sort order in terms of an ordered list of Indological transliteration units, where some transliteration units correspond to a single Brāhmī Unicode character (e.g., *ḥ* = U+11047 ʙʀᴀʜᴍɪ sɪɢɴ ᴠɪsᴀʀɢᴀ); some to a particular sequence of Brāhmī Unicode characters (e.g., *k* = U+11010 ʙʀᴀʜᴍɪ ʟᴇᴛᴛᴇʀ ᴋᴀ + U+1104D ʙʀᴀʜᴍɪ sɪɢɴ ᴠɪʀᴀᴍᴀ); some to either one of two alternate Brāhmī Unicode characters (e.g., *o* = either U+1100C ʙʀᴀʜᴍɪ ʟᴇᴛᴛᴇʀ ᴏ or U+11040 ʙʀᴀʜᴍɪ ᴠᴏᴡᴇʟ sɪɢɴ ᴏ) and some to part of one or more Brāhmī Unicode characters (e.g., Tocharian *ä* = either the inherent vowel of U+1108A ʙʀᴀʜᴍɪ ʟᴇᴛᴛᴇʀ ᴄᴇɴᴛʀᴀʟ ᴀsɪᴀɴ ᴋᴀ etc., or the inherent vowel of U+11029 ʙʀᴀʜᴍɪ ʟᴇᴛᴛᴇʀ ʏᴀ etc. + U+11095 ʙʀᴀʜᴍɪ sɪɢɴ ᴄᴇɴᴛʀᴀʟ ᴀsɪᴀɴ ᴅᴏᴜʙʟᴇ ᴅᴏᴛ). Please compare the descriptions of the individual writing systems above, and the transliterations given in the right-hand column of the character name list below. Note that some transliteration units have more than one meaning (namely, *ḷ* in Sanskrit, and *ḷ*, *r̠*, *ṉ* in Tamil vs. Tocharian and Uigur), and that two different units can be used to transliterate the same character in different languages (Uigur *δa* = Tumshuqese Saka *ḏa*). Also note that when *ṃ* is immediately followed by a stop, it is pronounced and sorted like the nasal consonant homorganic with that stop (e.g., like *ṅ* when followed by *k*, *kh*, *g*, *gh*, or *ṅ*). The implementation of this sorting algorithm in terms of Brāhmī Unicode character sequences directly is nontrivial and beyond our technical skills, but we shall be happy to evaluate the adequacy of implementations developed by others.

Brāhmī sort order: *a, ā, ä, ǟ, i, ī, u, ū, ü, ǖ, ṛ, ṝ, ḷ* (vocalic), *ḹ, e, ai, o, ö, ő, au, ṃ, ḥ, ẖ, ḫ, k, ḳ, kh, g, gh, ṅ, c, ch, j, jh, ñ, ṭ, ṭh, ḍ, ḍh, ṇ, t, ṯ, th, d, dh, n, ṉ* (Tocharian and Uigur), *p, ph, b, bh, m, ṃ, y, r, r̠* (Tocharian and Uigur), *l, ḷ* (Tocharian and Uigur), *v, w, ś, ś́, ṣ, ṣ̱, s, ṣ, h, ḻ* (Tamil), *ḷ* (consonantal), *r̠* (Tamil), *ṉ* (Tamil), *q, γ, δ / ḏ, dz, z, ź, χ š.*

9. Note on Vedic characters

Accent marks and other Vedic special characters have been excluded from this proposal because they are no more closely associated with pre-modern Brāhmī than with the modern

Brāhmī-derived scripts, and indeed the oldest extant Vedic manuscripts do not bear any accent marking. A block of Vedic characters, combinable both with Brāhmī and with the modern scripts, will need to be proposed separately.

Acknowledgements

Bibliography

Baums, Stefan, 2002. Jyotiṣkāvadāna. In: Jens Braarvig, ed., *Buddhist manuscripts*, volume II, Oslo: Hermes Publishing, 2002 (Manuscripts in the Schøyen Collection, III), pp. 287–302.

Bühler, G., 1894. The Bhattiprolu inscriptions. *Epigraphia Indica: a collection of inscriptions supplementary to the Corpus Inscriptionum Indicarum of the Archæological Survey*, II, pp. 323—329.

Bühler, G., 1896. *Indische Palaeographie von circa 350 a. Chr. – circa 1300 p. Chr.* Strassburg: Verlag von Karl J. Trübner. (Grundriss der indo-arischen Philologie und Altertumskunde, I. Band, 11. Heft

Dani, Ahmad Hasan, 1986. *Indian palaeography.* Second edition. New Delhi: Munshiram Manoharlal Publishers.

Hitch, Doug, 1981. *Central Asian Brahmi palaeography: the relationships among the Tocharian, Khotanese, and Old Turkic Gupta scripts.* MA thesis, Department of Linguistics, University of Calgary.

Konow, Sten, 1935. Ein neuer Saka-Dialekt. *Sitzungsberichte der Preußischen Akademie der Wissenschaften, philosophisch-historische Klasse*, pp. 772–823.

Konow, Sten, 1947. The oldest dialect of Khotanese Saka. *Norsk tidsskrift for sprogvidenskap*, XIV, pp. 156–190.

Lüders, Heinrich, 1912. Epigraphische Beiträge. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften*, pp. 806ff.

Mahadevan, Iravatham, 2003. *Early Tamil epigraphy: from the earliest times to the sixth century A.D.* Chennai, India: Cre-A:. (Harvard Oriental Series, volume sixty-two.)

Salomon, Richard, 1998. *Indian epigraphy: a guide to the study of inscriptions in Sanskrit, Prakrit, and the other Indo-Aryan languages.* New York: Oxford University Press. (South Asia Research.)

Sander, Lore, 1968. *Paläographisches zu den Sanskrithandschriften der Berliner Turfansammlung.* Wiesbaden: Franz Steiner Verlag. (Verzeichnis der orientalischen Handschriften in Deutschland, Supplementband 8.)

Sander, Lore, 1986. Brāhmī scripts on the Eastern Silk Roads. *Studien zur Indologie und Iranistik*, 11/12, pp. 159–192.

von Gabain, A., 1950. *Alttürkische Grammatik: mit Bibliographie, Lesestücken und Wörterverzeichnis, auch Neutürkisch.* 2. verbesserte Auflage. Leipzig: Otto Harrassowitz. (Porta linguarum Orientalium: Sammlung von Lehrbüchern für das Studium der orientalischen Sprachen, XXIII.)

Appendix: Samples of Brāhmī texts
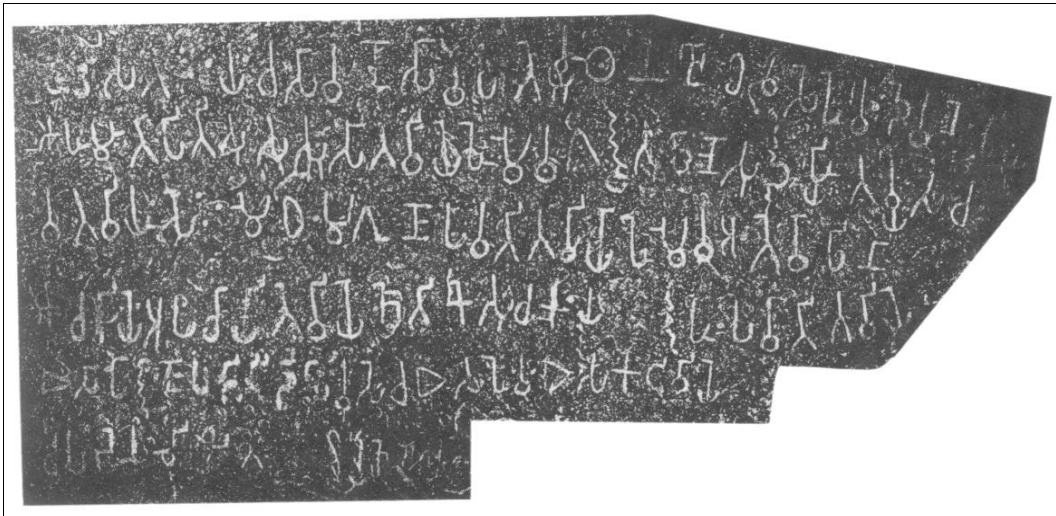
Illustration 1: Minor Rock Edict of Aśoka at Brahmagiri (Mahadevan 2003, p. 174).
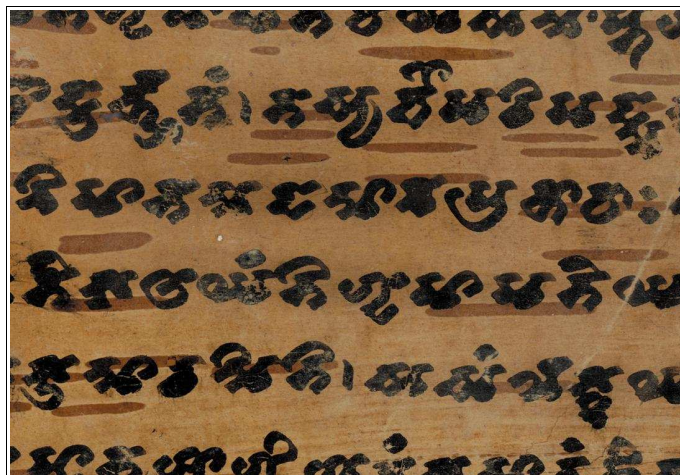
Illustration 2: Manuscript of the Jyotiṣkāvadāna in Gilgit-Bamiyan type I Brāhmī (Baums 2003, plate XVI.1).
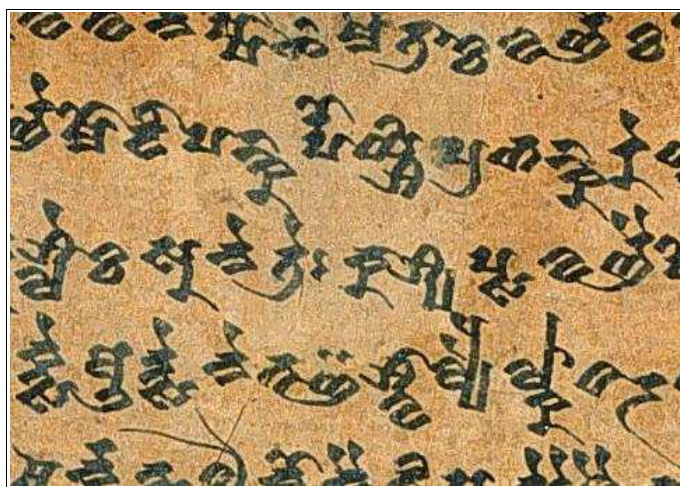
Illustration 3: Tocharian manuscript from Shorchuq (Staatsbibliothek zu Berlin).

u in such an early type of Central
uq Saka manuscripts only, the Pell
uq region seems to be the oldest.
mī, type a". ṇa has a tail ,
shape in most of the manuscripts,
ter than ta , which has
script the Tocharian alphabet only w
signs[4] had not been introduced
ment du koutchéen".). Moreover,

Illustration 4: Example of Brāhmī characters in modern
scholarly use (Sander 1986, p. 165).

	1100	1101	1102	1103	1104	1105	1106	1107
0	11000	11010	11020	11030	11040	11050	11060	11070
1	11001	11011	11021	11031	11041	11051	11061	11071
2	11002	11012	11022			11052	11062	11072
3	11003	11013	11023			11053	11063	11073
4	11004	11014	11024			11054	11064	11074
5	11005	11015	11025	11035	11045	11055	11065	11075
6	11006	11016	11026	11036		11056		11076
7	$(\bar{r})$ 11007	11017	11027	11037	11047	11057	11067	
8	(l) 11008	11018	11028	11038	11048	11058	11068	
9	$(\bar{l})$ 11009	11019	11029	11039	11049	11059		
A	1100A	1101A	1102A	1103A		1105A		
B	1100B	1101B	1102B	1103B		1105B		
C	1100C	1101C	1102C	$(-l)$ 1103C		1105C		
D	1100D	1101D	1102D	$(-\bar{l})$ 1103D	1104D	1105D		
E		1101E	1102E	1103E		1105E		
F		1101F	1102F	1103F		1105F		

16

	1108	1109	110A	110B	110C	110D	110E	110F
0	ၐ 11080	ၐ 11090						
1	ၐ 11081	ၐ 11091						
2	ၐ 11082	ၐ 11092	ၐ 110A2					
3		ၐ 11093						
4		ၐ 11094						
5		ၐ 11095						
6	ၐ 11086							
7								
8								
9		ၐ 11099						
A	ၐ 1108A	ၐ 1109A						
B	ၐ 1108B	ၐ 1109B						
C	ၐ 1108C	ၐ 1109C						
D	ၐ 1108D	ၐ 1109D						
E	ၐ 1108E	ၐ 1109E						
F	ၐ 1108F							

Independent vowel signs

11000	꣩	BRAHMI LETTER A	*a*
11001	꣩	BRAHMI LETTER AA	*ā*
11002	꣩	BRAHMI LETTER I	*i*
11003	꣩	BRAHMI LETTER II	*ī*
11004	꣩	BRAHMI LETTER U	*u*
11005	꣩	BRAHMI LETTER UU	*ū*
11006	꣩	BRAHMI LETTER VOCALIC R	*ṛ*
11007	(*r̄*)	BRAHMI LETTER VOCALIC RR	*ṝ*
11008	(*l̥*)	BRAHMI LETTER VOCALIC L	*ḷ*
11009	(*l̄*)	BRAHMI LETTER VOCALIC LL	*ḹ*
1100A	꣩	BRAHMI LETTER E	*e*
1100B	꣩	BRAHMI LETTER AI	*ai*
1100C	꣩	BRAHMI LETTER O	*o*
1100D	꣩	BRAHMI LETTER AU	*au*
1100E		<reserved>	
1100F		<reserved>	

Consonants

11010	꣩	BRAHMI LETTER KA	*ka*
11011	꣩	BRAHMI LETTER KHA	*kha*
11012	꣩	BRAHMI LETTER GA	*ga*
11013	꣩	BRAHMI LETTER GHA	*gha*
11014	꣩	BRAHMI LETTER NGA	*ṅa*
11015	꣩	BRAHMI LETTER CA	*ca*
11016	꣩	BRAHMI LETTER CHA	*cha*
11017	꣩	BRAHMI LETTER JA	*ja*
11018	꣩	BRAHMI LETTER JHA	*jha*
11019	꣩	BRAHMI LETTER NYA	*ña*
1101A	꣩	BRAHMI LETTER TTA	*ṭa*
1101B	꣩	BRAHMI LETTER TTHA	*ṭha*
1101C	꣩	BRAHMI LETTER DDA	*ḍa*
1101D	꣩	BRAHMI LETTER DDHA	*ḍha*
1101E	꣩	BRAHMI LETTER NNA	*ṇa*
1101F	꣩	BRAHMI LETTER TA	*ta*
11020	꣩	BRAHMI LETTER THA	*tha*
11021	꣩	BRAHMI LETTER DA	*da*
11022	꣩	BRAHMI LETTER DHA	*dha*
11023	꣩	BRAHMI LETTER NA	*na*
11024	꣩	BRAHMI LETTER PA	*pa*
11025	꣩	BRAHMI LETTER PHA	*pha*
11026	꣩	BRAHMI LETTER BA	*ba*
11027	꣩	BRAHMI LETTER BHA	*bha*
11028	꣩	BRAHMI LETTER MA	*ma*

11029	౿	BRAHMI LETTER YA	*ya*
1102A	౿	BRAHMI LETTER RA	*ra*
1102B	౿	BRAHMI LETTER LA	*la*
1102C	౿	BRAHMI LETTER VA	*va*
1102D	౿	BRAHMI LETTER SHA	*śa*
1102E	౿	BRAHMI LETTER SSA	*ṣa*
1102F	౿	BRAHMI LETTER SA	*sa*
11030	౿	BRAHMI LETTER HA	*ha*
11031	౿	BRAHMI LETTER LLA	*ḷa*
11032		<reserved>	
11033		<reserved>	
11034		<reserved>	

Dependent vowel signs

11035	౿	BRAHMI VOWEL SIGN AA	*ā*
11036	౿	BRAHMI VOWEL SIGN I	*i*
11037	౿	BRAHMI VOWEL SIGN II	*ī*
11038	౿	BRAHMI VOWEL SIGN U	*u*
11039	౿	BRAHMI VOWEL SIGN UU	*ū*
1103A	౿	BRAHMI VOWEL SIGN VOCALIC R	*ū*
1103B	౿	BRAHMI VOWEL SIGN VOCALIC RR	*r̄*
1103C	(-*l̥*)	BRAHMI VOWEL SIGN VOCALIC L	*ḷ*
1103D	(-*l̄*)	BRAHMI VOWEL SIGN VOCALIC LL	*ḹ*
1103E	౿	BRAHMI VOWEL SIGN E	*e*
1103F	౿	BRAHMI VOWEL SIGN AI	*ai*
11040	౿	BRAHMI VOWEL SIGN O	*o*
11041	౿	BRAHMI VOWEL SIGN AU	*au*
11042		<reserved>	
11043		<reserved>	
11044		<reserved>	

Various signs

11045	౿	BRAHMI SIGN ANUSVARA	*ṃ*
11047	౿	BRAHMI SIGN VISARGA	*ḥ*
11048	౿	BRAHMI LETTER JIHVAMULIYA	*ẖa*
11049	౿	BRAHMI LETTER UPADHMANIYA	*ḫa*
1104D	౿	BRAHMI SIGN VIRAMA	
1104E		<reserved>	
1104F		<reserved>	

Numbers

| 11050 | ౿ | BRAHMI DIGIT ZERO | *0* |
| 11051 | ౿ | BRAHMI DIGIT ONE | *1* |

11052	ꠒ	BRAHMI DIGIT TWO	*2*
11053	ꠓ	BRAHMI DIGIT THREE	*3*
11054	ꠔ	BRAHMI DIGIT FOUR	*4*
11055	ꠕ	BRAHMI DIGIT FIVE	*5*
11056	ꠖ	BRAHMI DIGIT SIX	*6*
11057	ꠗ	BRAHMI DIGIT SEVEN	*7*
11058	ꠘ	BRAHMI DIGIT EIGHT	*8*
11059	ꠙ	BRAHMI DIGIT NINE	*9*
1105A	ꠚ	BRAHMI NUMBER TEN	*10*
1105B	ꠛ	BRAHMI NUMBER TWENTY	*20*
1105C	ꠜ	BRAHMI NUMBER THIRTY	*30*
1105D	ꠝ	BRAHMI NUMBER FOURTY	*40*
1105E	ꠞ	BRAHMI NUMBER FIFTY	*50*
1105F	ꠟ	BRAHMI NUMBER SIXTY	*60*
11060	ꠠ	BRAHMI NUMBER SEVENTY	*70*
11061	ꠡ	BRAHMI NUMBER EIGHTY	*80*
11062	ꠢ	BRAHMI NUMBER NINETY	*90*
11063	ꠣ	BRAHMI NUMBER ONE HUNDRED	*100*
11064	ꠤ	BRAHMI NUMBER TWO HUNDRED	*200*
11065	ꠥ	BRAHMI NUMBER THREE HUNDRED	*300*
11066		<reserved>	
11067	ꠧ	BRAHMI NUMBER ONE THOUSAND	*1000*
11068	꠨	BRAHMI NUMBER TWO THOUSAND	*2000*
11069		<reserved>	
1106A		<reserved>	
1106B		<reserved>	
1106C		<reserved>	
1106D		<reserved>	
1106E		<reserved>	
1106F		<reserved>	

Punctuation

11070	꠰	BRAHMI DANDA	|
11071	꠱	BRAHMI DOUBLE DANDA	‖
11072	꠲	BRAHMI PUNCTUATION DOT	·
11073	꠳	BRAHMI PUNCTUATION DOUBLE DOT	:
11074	꠴	BRAHMI PUNCTUATION LINE	-
11075	꠵	BRAHMI PUNCTUATION CRESCENT BAR	€
11076	꠶	BRAHMI PUNCTUATION LOTUS	❁
11077		<reserved>	
11078		<reserved>	
11079		<reserved>	
1107A		<reserved>	
1107B		<reserved>	
1107C		<reserved>	
1107D		<reserved>	

| 1107E | | <reserved> | |
| 1107F | | <reserved> | |

Tamil Brāhmī signs

11080	ഴ	BRAHMI LETTER TAMIL LLLA	*ḻa*
11081	൬	BRAHMI LETTER TAMIL RRA	*ṟa*
11082	௴	BRAHMI LETTER TAMIL NNA	*ṉa*
11083		<reserved>	
11084		<reserved>	
11085		<reserved>	

Bhattiprolu Brāhmī sign

11086	᷏	BRAHMI VOWEL SIGN BHATTIPROLU AAA	*ā*
11087		<reserved>	
11088		<reserved>	
11089		<reserved>	

Central Asian Brāhmī signs

1108A	ଃ	BRAHMI LETTER CENTRAL ASIAN KA	*ḵa*
1108B	ଢ	BRAHMI LETTER CENTRAL ASIAN TA	*ṭa*
1108C	ଡ	BRAHMI LETTER CENTRAL ASIAN NA	*ṉa*
1108D	ଗ	BRAHMI LETTER CENTRAL ASIAN PA	*pa*
1108E	ଖ	BRAHMI LETTER CENTRAL ASIAN MA	*ṃa*
1108F	ଌ	BRAHMI LETTER CENTRAL ASIAN RA	*ṟa*
11090	୭	BRAHMI LETTER CENTRAL ASIAN LA	*ḻa*
11091	ଈ	BRAHMI LETTER CENTRAL ASIAN SHA	*śa*
11092	ଙ	BRAHMI LETTER CENTRAL ASIAN SSA	*ṣa*
11093	ଫ	BRAHMI LETTER CENTRAL ASIAN SA	*ṣa*
11094	ଵ	BRAHMI LETTER CENTRAL ASIAN WA	*wa*
11095	்	BRAHMI SIGN CENTRAL ASIAN DOUBLE DOT	*ä*
11096		<reserved>	
11097		<reserved>	
11098		<reserved>	
11099	ଃ	BRAHMI LETTER CENTRAL ASIAN QA	*qa*
1109A	ଃ	BRAHMI LETTER CENTRAL ASIAN GA	*γa*
1109B	௸	BRAHMI LETTER CENTRAL ASIAN DA	*δa* / *ḏa*
1109C	ଃ	BRAHMI LETTER CENTRAL ASIAN DZA	*dᶻa*
1109D	ଃ	BRAHMI LETTER CENTRAL ASIAN ZA	*za*
1109E	௵	BRAHMI LETTER CENTRAL ASIAN ZHA	*źa*
1109F		<reserved>	
110A0		<reserved>	
110A1		<reserved>	
110A2	ୀ	BRAHMI LETTER CENTRAL ASIAN KHSHA	*χša*
110A3		<reserved>	

110A4	<reserved>
110A5	<reserved>
110A6	<reserved>
110A7	<reserved>
110A8	<reserved>
110A9	<reserved>
110AA	<reserved>
110AB	<reserved>
110AC	<reserved>
110AD	<reserved>
110AE	<reserved>
110AF	<reserved>
110B0	<reserved>
...	
110FF	<reserved>