# **9** Writing systems

Edited by Elena Bashir

#### 9.1. Introduction

By Elena Bashir

The relations between spoken language and the visual symbols (graphemes) used to represent it are complex. Orthographies can be thought of as situated on a continuum from "deep" — systems in which there is not a one-to-one correspondence between the sounds of the language and its graphemes — to "shallow" — systems in which the relationship between sounds and graphemes is regular and transparent (see Roberts & Joyce 2012 for a recent discussion). In orthographies for Indo-Aryan and Iranian languages based on the Arabic script and writing system, the retention of historical spellings for words of Arabic or Persian origin increases the orthographic depth of these systems. Decisions on how to write a language always carry historical, cultural, and political meaning. Debates about orthography usually focus on such issues rather than on linguistic analysis; this can be seen in Pakistan, for example, in discussions regarding orthography for Kalasha, Wakhi, or Balti, and in Afghanistan regarding Wakhi or Pashai. Questions of orthography are intertwined with language ideology, language planning activities, and goals like literacy or standardization. Woolard 1998, Brandt 2014, and Sebba 2007 are valuable treatments of such issues.

In Section 9.2, Stefan Baums discusses the historical development and general characteristics of the (non Perso-Arabic) writing systems used for South Asian languages, and his Section 9.3 deals with recent research on alphasyllabic writing systems, script-related literacy and language-learning studies, representation of South Asian languages in Unicode, and recent debates about the Indus Valley inscriptions. Elena Bashir's Section 9.4 treats adaptations of the Perso-Arabic script used for various languages of South Asia, and her Section 9.5, on current research areas and desiderata, concludes the chapter.

# 9.2. General historical and analytical By Stefan Baums

# 9.2.1. Early scripts

We have no securely datable documents from South Asia until the edicts of the emperor Asoka in the third century BCE.<sup>1</sup> The edicts, throughout Asoka's empire in northern and parts of southern India, are written in two different scripts: those in the northwest (modern Afghanistan and Pakistan) in Kharosthī, the rest in Brāhmī. Greek and Aramaic versions of the edicts have likewise been found in Afghanistan. The connections of the Kharosthī script and scribal institutions with the Achaemenid Empire suggest that Kharosthī was invented not later than the fourth century BCE (Baums 2014) and that Asoka was following established custom when he used Kharosthi in his northwestern edicts. The Sanskrit grammar of Pānini (fourth century BCE and a native of the northwest) likewise mentions writing (lipi) and books (grantha), but it remains unclear whether he referred to Aramaic or Kharosthī script. Around 325 BCE, Alexander's general Nearchos observed that the inhabitants of the northwest wrote letters on tightly woven cloth (whether in Aramaic or Kharosthī). Some twenty years later, the Greek ambassador Megasthenes denied that writing was used for legal proceedings in northeastern India, but this does not preclude its use for other purposes (v. Hinüber 1990). The evidence thus indicates a continuous writing tradition since Achaemenid times in northwestern South Asia, with a transition from Aramaic to Kharosthī at some point before the third century BCE. In contrast, there is no conclusive evidence for writing in mainland India until the time of Asoka, and it remains possible, though by no means certain, that he or his immediate predecessors may have invented the Brāhmī script on the model of Kharosthī.

# 9.2.1.2. Kharosthī

Kharoṣṭhī remained throughout its existence a regional script of the northwest (ancient Gandhāra). Following the Silk Roads, it spread into Central Asia and in the third century CE was used to write administrative and legal records in the area of Loulan and Niya in the southern Tarim basin. Around the same time, expatriate South Asian Buddhist communities used Kharoṣṭhī in the capital of China. In its homeland, Kharoṣṭhī died out by the fourth century CE and was replaced by variants of Brāhmī; in the northern Tarim basin it continued to be used for another century or two.

<sup>&</sup>lt;sup>1</sup> The discussion in this section is an updated and expanded version of Baums 2011: 240-250.

Like its model, the Aramaic script used by the Achaemenid administration in Gandhāra, Kharoṣṭhī is written from right to left, and the two scripts agree in the shapes and sound values of some (but by no means all) of their letters. While Aramaic writes only the consonants of words, not their vowels, the developers of Kharoṣṭhī added an original system of marking vowels that became the model for Brāhmī and all later scripts derived from it. A consonant letter without modification, such as  $\rightarrow$ , signifies not just the bare consonant, in this case k, but the syllable ka – the vowel a is said to be "inherent" in consonant signs. If a vowel other than a follows a consonant, its presence is indicated by attaching a vowel diacritic to the consonant sign:  $\rightarrow$ , e.g., is ki, and  $\rightarrow$  is ku. The phonemic difference between short and long vowels is not marked:  $\rightarrow$  is used to write both [ki] and [ki:], etc. Some scribes indicate preconsonant lanasal segments by a hook (anusvāra, transliterated m) under the preceding consonant sign: compare  $\rightarrow ka$  with p kam.

Where one consonant immediately follows another without intervening vowel, this is indicated by combining the two consonant signs into one conjunct consonant or ligature: the sequence *tsa*, for instance, is represented by combining the signs for *ta*, >, and *sa*, ?, into the ligature .<sup>3</sup>, thus cancelling the vowel *a* inherent in *t*; the sequence *tasa*, on the other hand, is written ?>. While most consonant clusters are written with transparent conjuncts, some, such as 3 stam or y ksa, are indicated by opaque (and possibly atomic) signs. In order to write a vowel at the beginning of a word or a vowel following another vowel, its vowel diacritic is attached to the "vowel carrier" ?; ? alone signifies *a*, ? is *i*, 3 is *u*, and so on. (It is likely that this so-called vowel carrier was in reality the consonant [?].) The graphical unit of a simple consonant sign or ligature followed by an optional vowel diacritic and/or anusvāra is called an akṣara. It forms the basic graphical unit of Kharoṣṭhī and all Brāhmī-derived scripts.

The letters of the Kharoṣṭhī script are conventionally arranged in the alphabetic order *a*, *ra*, *pa*, *ca*, *na*, etc. — the so-called *arapacana* alphabet. Its origins are obscure, but it seems to have undergone systematic extension (Baums 2009: 194–197) and, after the demise of Kharoṣṭhī, lived on as a Buddhist magical formula (Brough 1977). Equally obscure is the origin of the name "Kharoṣṭhī" (see Falk 1993: 84–90 for a summary of theories). Kharoṣṭhī was almost exclusively used for the Middle Indo-Aryan literary language Gāndhārī, but attempts were made to adapt it to the writing of Sanskrit (Salomon 2008, Strauch 2012).

# 9.2.1.3. Brāhmī

In contrast to Kharoṣṭhī, the other early South Asian writing system Brāhmī is (almost always) written from left to right, and where Kharoṣṭhī (like Aramaic) has a distinctly cursive ductus, early Brāhmī is a monumental script whose letters consist of straight lines, circles, and other basic geometric shapes. Both properties may have been inspired by some degree of familiarity with the Greek script (Falk 1993:

109–112). The general system of Brāhmī is the same as that of Kharoṣṭhī, but it improves on the older script by using separate diacritics for short and long vowels; Brāhmī also employs independent signs for initial vowels instead of a vowel carrier plus diacritic like Kharoṣṭhī: ka,  $k\bar{a}$ , ki,  $k\bar{r}$  are written +, f, f, and syllable-initial a,  $\bar{a}$ , i,  $\bar{i}$  are written  $\aleph$ ,  $\Re$ ,  $\therefore$  Brāhmī, like Kharoṣṭhī, was originally developed for Middle Indo-Aryan and only later came to be used for Sanskrit. Among the earlier Brāhmī inscriptions, conjunct consonants are therefore rare, and where they occur in the Girnar edict of Aśoka they may represent coarticulations rather than true clusters (& pta, e.g., may be a labialized [t<sup>w</sup>] < OIA [tv]; v. Hinüber 2001: 196– 197). The name "Brāhmī" may refer to its use (starting in the first century CE) for writing the language of the Brahmans, Sanskrit (Falk 1993: 106–108).

The letters of the Brāhmī script are arranged in an alphabetic order called varnamālā ('sound garland') that is based on phonological principles. In the full form used for Sanskrit, first come the simple vowels in pairs of short and long, then the diphthongs; then anusvāra and visarga (h, syllable-final voiceless [h]); then the stop consonants in the order unvoiced plain and aspirate, voiced plain and aspirate, and nasal (in order of place of articulation, starting in the back of the mouth); then the four semivowels ya, ra, la, va; then the sibilants sa, sa, sa; and finally, ha (voiced [h]).

Early Brāhmī (third to first century BCE) became the ancestor of all later South Asian (as well as the main Southeast and some Central Asian) scripts. With few exceptions, the akṣara system remained constant, and only the forms of consonant signs, vowel signs, and vowel diacritics continued to evolve. Following Dani (1963), we can distinguish the following stages of development: Early, Middle, and Late Brāhmī.

# 9.2.1.3.1. Early Brāhmī

Changes in letter shape resulted from cursivization or stroke reduction (Aśokan + ka, e.g., turned into +, and  $\perp na$  into +); from changes in the order or direction of strokes (Aśokan  $\pm na$  became  $\times$  and later +); and from the incorporation into the writing system of originally insignificant mechanical changes. The last type of change had the greatest effect on the ductus of scripts. The little blot of ink, for instance, that a stylus leaves where it first touches the writing surface at the top of letters later developed into the long horizontal headlines of Nagari and Bengali and the variously shaped heads of the Southern scripts, such as the "check mark" of Telugu or the "umbrella" of Oriya.

Regional variations first appeared in the Early Brāhmī period (third to first centuries BCE). The letter b *dha* was mirrored to a, h *ga* was cursivized to a and b *ma* was angularized to  $\mathbf{x}$ . More radical systemic changes occurred in Old Tamil cave inscriptions of the second and first centuries BCE. Mahadevan's (2003) system TB-I uses the vowel mātrā  $\bar{a}$  to represent both short and long [a] / [a:]; in this orthography consonant signs without the  $\bar{a}$  mātrā do not have an inherent vowel and always represent the bare consonant. In system TB-II, the  $\bar{a}$  mātrā always represents long [a:], whereas vowelless consonant signs can be read either with inherent short [a] or as bare consonants. By the second century CE, the ambiguity of  $\bar{a}$  in TB-I and of bare consonant signs in TB-II (and the influence of the standard Brāhmī system) led to system TB-III in which a dot (*pulli*) marks the absence of a vowel when placed above basic consonant signs, and shortening when used with the vowels *e* and *o* (in contrast to Old and Middle Indo-Aryan, Dravidian languages have short as well as long *e* and *o* phonemes). Old Tamil Brāhmī added signs for *l*, *l*, *r*, *n* and possibly *n* to the original inventory of Brāhmī.

Middle Indo-Aryan inscriptions from Bhattiprolu in South-East India (second century BCE) employ two separate diacritics for short and long a ( $\mathbf{f}$  is ka,  $\mathbf{f}$  is  $k\bar{a}$ ), which seems to be an extension of the Old Tamil Brāhmī system TB-I. The main reason for abandoning inherent [a] in Tamil Brāhmī does not apply in the case of the Bhattiprolu inscriptions since Middle Indo-Aryan does not have word-final consonants or non-homorganic clusters. This implies that the dedicated long  $\bar{a}$  mātrā, too, was first introduced in a Tamil context, and the resulting system only later imitated in Bhattiprolu, but no such Tamil inscription has yet been discovered.

# 9.2.1.3.2. Middle Brāhmī

In the Middle Brāhmī period (first to third centuries CE), local variants became more distinct; Dani (1963) distinguishes Kauśāmbī, Mathurā, Western Deccan, and Eastern Deccan regional styles. Headmarks developed different shapes (linear, square, triangular, etc.), there were further angularizations ( $\varepsilon$  *ja*) and cursivizations ( $\varkappa$  *sa*), and the vowel diacritics tended to assume more elaborate forms (as in  $2 d\overline{i}$ ). In this period, Sanskrit was first used in inscriptions and manuscripts, and additional signs were introduced to represent its sounds ( $\varepsilon$  *kr*,  $\ast$  *kau*,  $\varepsilon$  *kah*,  $\varepsilon$ *na*). Northwestern manuscripts of the first century CE contain the first examples of vowel cancellation marker (virāma) indicating cancellation of inherent *a* wherever there was no following consonant sign to form a ligature with. This early virāma device consists in lowering the sign for the vowelless consonant below the baseline, linking it with the preceding akṣara, and putting a short horizontal line on top of it (as in  $\frac{c}{2\pi} tv\bar{a}t$ ). In the later scripts, just the equivalents of that horizontal line, now placed diagonally below the consonant sign, are used as virāma, with the consonant sign in normal position (cf. Devanāgarī calc  $tv\bar{a}t$ ).

# 9.2.1.3.3. Late Brāhmī

In the Late Brāhmī period (fourth to sixth centuries CE), graphical differentiation reached a point where regional forms of Brāhmī would have to be learned separately and we can therefore speak of different scripts rather than variants of a single script. While older scholarship distinguished between "Western" and "Eastern" Gupta scripts and South Indian Brāhmī, Dani (1963) suggests a categorization into nine main geographical divisions. In the South Indian scripts, letters began to assume their typical round forms because they were now incised into the surface of palm leaves, then rubbed with ink, and straight lines would have tended to rupture the leaf. Some letters reached their modern forms (compare e.g. northern 9 ga with Nāgarī  $\overline{\textbf{n}}$ ). In Central Asia, local scripts were developed on the basis of a northwestern Gupta type, in order to write Sanskrit locally and for the writing of Tocharian, Uyghur, and Khotanese.

# 9.2.2. Transitional Script period

In the Transitional Script period (seventh to tenth centuries CE), proto-Śāradā ("Gilgit-Bamiyan type II" in Sander's 1968 terminology) emerged as a distinct northwestern script. In its fully developed form, Śāradā was used for the writing of Sanskrit and Kashmiri and gave rise to the regional traders' scripts Takri (used for Western Pahari) and Landa (used for Sindhi and Panjabi). In the rest of Northern India, the Siddhamātrkā script was in use, eventually giving rise to Nagari and Bengali and living on in East Asia as the "Siddham" script. The Tibetan writing system was developed at the beginning of the Transitional Script period under Central Asian and North Indian influences (van Schaik 2011 argues for a predominance of the latter). In upper South India, a distinct proto-Kannada-Telugu script began to take shape. In the far South, three different scripts were emerging: the Grantha script for Sanskrit, and the Tamil and cursive Vaṭteluttu scripts for Tamil. The Grantha script used by the Pallava dynasty became the basis of the Southeast Asian scripts. The Sinhalese script had so far mostly developed in isolation; now it was subjected to a strong influence from Pallava Grantha.

# 9.2.3. Modern Devanagari and Gujarati

The modern Nagari (or Devanagari) script (Maurer 1976) had assumed a distinct identity by the 11th century and is now used throughout northern India for Hindi, Nepali, Marathi, local dialects like Bhojpuri, and non-Indo-Aryan languages like Gondi.<sup>2</sup> From the 11th to 16th centuries, a regional form called Nandinagari was used in southern India, and between the 12th and 16th centuries, an ornamental variant called Ranjana took shape in Nepal under influence from Bengal. Sanskrit is primarily printed in Nagari today, and in spite of strong loyalty to the local scripts, Nagari has developed a presence throughout modern India.

<sup>&</sup>lt;sup>2</sup> Some of the examples for the modern Indic scripts are drawn from Bright & Daniels 1996.

Nagari letters are characterized by horizontal headlines and right angles. The vowel diacritics for  $\bar{a}$ , i,  $\bar{i}$ , and o have drooped to the base line of the akṣara (in the case of i on its left side): compare  $\bar{a}\pi k\bar{a}$  and  $\bar{b}\kappa ki$  with their remote ancestors in Early Brāhmī,  $\bar{f}$  and f. At a stage in the linguistic prehistory of Hindi, word-final, and under certain conditions word-medial, short [ə] disappeared, but this change is not mirrored by the writing system of Hindi, so that 'Monday', e.g., is written  $\pi i H \pi i x somav \bar{a}ra$  but pronounced [so:mva:r]. For the representation of peripheral phonemes that have entered Hindi via loanwords, a subscript dot (nuqtā) is optionally added to consonant signs of similar pronunciation:  $\bar{a}r qa$ ,  $\bar{a}r xa$ ,  $\bar{a}r ya$ ,  $\bar{a}r za$ , and  $\bar{a}r fa$ , imitating the way the Arabic script was extended for the writing of languages such as Persian. Nagari gave rise to regional traders' scripts such as Modi (used for Marathi).

The earliest inscriptions in the Gujarati language, dating from the 15th century, are written in Nagari script. Later a cursive variant of Nagari began to develop into a separate Gujarati script, which only attained widespread currency in the middle of the 19th century. Gujarati betrays its cursive origin in the lack of a headline (compare  $\Im$  ga and  $\dashv$  ta with their Nagari counterparts  $\overline{\neg}$  and  $\overline{\neg}$ ) and developed a consistently analytic notation for initial vowels: the signs for  $\overline{a}$  ( $\Im$ ), e ( $\eth$ ), ai ( $\eth$ ), o ( $\Im$ ), and au ( $\Im$ ) are all derived by diacritic vowel signs from the sign for initial a ( $\Im$ ), while in Nagari this is only true of  $\overline{a}$  ( $\Im$ ), o ( $\Im$ ), and au ( $\Im$ ) (all from  $\Im$ ).

# 9.2.4. Gurmukhi and Khojki

The Gurmukhi script is used for Panjabi, especially by Sikhs in and from the Indian Panjab. It was developed by Guru Angad (1504-1552) on the background of Landa and takes its name from this fact and from its use in the Adi Granth. Gurmukhi has a unique system of writing initial vowels by adding the vowel diacritics to one of three different vowel carriers:  $\mathfrak{M}$  is used for  $\mathfrak{M} a$ ,  $\mathfrak{M} \overline{a}$ ,  $\mathfrak{N} ai$ , and ਔ au (low vowels and diphthongs); ੲ for ਇ i, ਈ ī, and ਏ e (front vowels); and ੳ for  $\Theta u$ ,  $\Theta \bar{u}$ , and  $\Theta o$  (back vowels). Panjabi lost the voiced aspirates and developed a system of high, mid, and low tones. Synchronically, the "voiced aspirate" letters have the following tone-marking functions: a vowel preceded by a "voiced aspirate" (stem-initial, or stem-medial between a short and a long vowel) carries a low tone (ਘੋੜਾ ghorā [kora], ਪੰਘਾਰਨਾ paghāranā [pəgàrna]); a vowel followed by a stem-final "voiced aspirate" letter carries a high tone (HT4 māgha [mág]); non-initial ਹ ha also represents high tone on the preceding vowel (ਤੀਹ tīha [tí]). Another innovation is a diacritic called addak, marking gemination of consonants (ਪॅवी pakkī), imitating the Arabic tašdīd. As in Nagari, peripheral loanword phonemes (and more recently, indigenous retroflex *l*) are marked by subscript dots.

Around the same time as the development of Gurmukhi, the Landa script was also adapted by the South Asian Ismaili community for the writing of their religious literature, the *ginān* (Asani 1987). The development of this Ismaili Khojki script

(from Persian *xwājah* 'master') is attributed to Pir Sadr al-Din (15th century). The main improvements of Khojki over the traders' script Landa consist in the addition of medial vowel marks (here called lākanā), of a gemination marker called šadda (corresponding to the Gurmukhi addak), and especially the introduction of systematic punctuation to separate words. The Khojki script fell out of general use by the 1970s and has been replaced by the Gujarati and Arabic scripts.

# 9.2.5. Bengali and Oriya

Like Nagari, the Bengali script derives from the north Indian Siddhamātrkā. The close relationship of Bengali script and Nagari is apparent from their use of a horizontal head line and the shape of letters such as  $\overline{\Phi}$  /  $\overline{\Phi}$  ka and  $\overline{\Phi}$  /  $\overline{\Phi}$  na. The ductus of the Bengali script is defined by acute angles. For the representation of postconsonantal e, ai, o, and au the Bengali script employs so-called prsthamātrās ('backstrokes'): e and ai are written to the left of the consonant sign (( $\overline{\Phi}$ ,  $\overline{\Phi}$ ), o and au surround it ((का, (को); the other modern scripts with prsthamātrās are Oriya, Malayalam, Tamil, and Sinhalese. While the orthography of most South Asian scripts is close to their pronunciation, Bengali orthography is very conservative and does not reflect many sound changes in the history of Bengali: The distinction between long and short  $i / \bar{i}$  and  $u / \bar{u}$  is only made in writing ( $\overline{\Phi q}$  kula and  $\overline{\Phi q}$  kula, e.g. are both pronounced [kul]); the three sibilants  $\Im$  sa,  $\exists$  sa, and  $\exists$  sa are pronounced the same (mostly [[], [s] before dentals);  $\P$  *na* and  $\overline{\P}$  *na* are both pronounced [n]. Conversely, there is only one sign  $(a, \overline{a})$  or inherent) for the not wholly predictable two pronunciations  $[\mathfrak{I}]$  and  $[\mathfrak{I}]$ ; and one ( $\mathfrak{L}$  or  $\mathfrak{I}$ ) for the pronunciations  $[\mathfrak{X}]$  and [e]. Original consonant clusters are written as such, but pronounced as geminates if the first element is medial, and as the simple first element if it is initial (初环 sbāsa [[a[], বিদ্বান bidbāna [biddan]). Nasal stops and y as second elements of clusters lead to nasalization and palatalization of following vowels (স্মারক smāraka [[ãrok], ব্যাকরণ byākaraņa [bækoron]).

The Oriya script is descended from proto-Bengali, but has been influenced by the ductus of South Indian scripts in the round shape of letters and in the "umbrella" covering most letters, corresponding to the head line of Nagari or Bengali (cf.  $\Im$  ka, Nagari  $\overline{\sigma}$ , Bengali  $\overline{\sigma}$ ;  $\Omega$ , Nagari  $\overline{\sigma}$ , Bengali  $\overline{\sigma}$ ).

# 9.2.6. South Indian scripts

Among the modern South Indian scripts, one subgroup is formed by the Kannada and Telugu scripts, another by the Grantha-derived Malayalam and Tamil. Kannada and Telugu have their first precursor in the script of the Kadamba and Cālukya inscriptions of the 5th to 7th centuries. After the 10th century, one can speak of a distinct Old Kannada script, which by the year 1500 had begun to differentiate into Kannada and Telugu varieties; the differences between the two scripts were standardized by their use in printing from the 19th century. The main distinction between the modern Kannada and Telugu scripts is the different shape of the headmark, which is a horizontal line with a hook at the right in Kannada (e.g. in  $\vec{\sigma} ka$ ,  $\vec{\omega} ca$ ,  $\vec{\sigma} ta$ ) but looks like a check mark on top of the letters in Telugu ( $\vec{s} ka$ ,  $\vec{\omega} ca$ ,  $\vec{\sigma} ta$ ). In both scripts, some of the aspirate letters are formed by addition of a diacritic subscript line or dot to the unaspirated letter (e.g. Telugu  $\vec{\omega} cha$  from  $\vec{\omega} ca$ , Kannada and Telugu  $\vec{\omega} / \vec{\omega} dha$  from  $\vec{\omega} / \vec{\omega} da$ ); the subscript line is then also applied to aspirates that have their own distinct letter (e.g. Kannada and Telugu  $\vec{\omega} / \vec{\omega} gha$ , cf.  $\vec{n} / \vec{\alpha} ga$ ). In Kannada only, the diacritics for the long vowels  $\bar{i}$ ,  $\bar{e}$ , and  $\bar{o}$ are derived from those of the corresponding short vowels:  $\vec{\vartheta} ki$  vs.  $\vartheta e k\bar{i}$ ,  $\vartheta ke$  vs.  $\vartheta e$  $k\bar{e}$ ,  $\vartheta ko$  vs.  $\vartheta e k\bar{o}$ . Also in Kannada, consonant clusters with initial *r* are written with the combining form of *r* following the other consonant:  $\vec{\vartheta} \in rta$ .

The Malayalam and Tamil script form another subgroup as shown by similarities of letter shape and systemic features. The vowel diacritic for short *i*, e.g., is placed to the right of the akşara only in these two scripts: Malayalam  $\mathfrak{G}$  *ki*, Tamil **G** *ki* vs. Kannada  $\mathfrak{F}$  *ki*, Telugu  $\mathfrak{F}$  *ki*. The vowel diacritics for  $\bar{a}$ , e,  $\bar{e}$ , ai, o,  $\bar{o}$ , and *au* are physically separated from the main part of the akşara. Malayalam and Tamil use pṛṣṭhamātrās for these vowels, again in contrast with Kannada and Telugu, but in common with Sinhalese (as well as Bengali and Oriya). Six Malayalam characters ( $\mathfrak{G}$  *ka*,  $\mathfrak{M}$  *na*,  $\mathfrak{O}$  *na*,  $\mathfrak{O}$  *ra*,  $\mathfrak{L}$  *la*, and  $\mathfrak{G}$  *la*) form ligatures with the virāma sign ( $\check{}$ ), the so-called cillakṣarams:  $\mathfrak{G}$ ,  $\mathfrak{M}$ ,  $\mathfrak{N}$ ,  $\mathfrak{N}$ ,  $\mathfrak{M}$  and  $\mathfrak{O}$  (cf.  $\mathfrak{G}$  *kh* or  $\mathfrak{O}$  *g*). Orthographic reforms in the 1970s and 1980s introduced new signs for postconsonantal *u*,  $\bar{u}$ , *r*, and *r* that are placed on the left and right side of the akṣara ( $\mathfrak{L}$  *pu*,  $\mathfrak{L}$  *pū*,  $\mathfrak{I}$  *pr*, and ( $\mathfrak{L}$  *pra*) and replaced all consonant conjuncts by combinations with virāma or cillakṣaram ( $\mathfrak{G}\mathfrak{O}$  *kta* became  $\mathfrak{G}$   $\mathfrak{O}$ ,  $\mathfrak{M}$  *nta* became  $\mathfrak{N}$   $\mathfrak{O}$ , etc.). The Malayalam script is primarily used for the Malayalam and Tulu languages.

The Tamil script can be traced back to the 9th century (like its cursive variant Vatteluttu) and assumed its modern form by the 15th century. The sign inventory of Tamil is much smaller than that of the other Brāhmī-derived scripts: signs for aspirate stops were abandoned because they do not occur in Tamil, and signs for voiced stops because they only occur as allophones of their voiced counterparts ( $\mathbf{5}$  ka represents both [k] and [g], etc.). There are separate vowel signs for short and long e and o. Tamil consistently uses the pulli sign instead of consonant conjuncts (with the exception of **Fop** ksa which is regarded as a basic letter). The character repertoire of Tamil has three layers: the core characters needed to write Tamil itself; five characters inherited from Grantha and used for Sanskrit words ( $\mathbf{2}$  ja,  $\mathbf{60}$  sa,  $\mathbf{60}$  sa,  $\mathbf{60}$  sa,  $\mathbf{60}$  ksa); and the visarga sign ( $\mathbf{..}$ ), called āytam. The last is also used in combination with other consonant signs to write peripheral loanword phonemes (corresponding functionally to the subscript dot, nuqtā, of Nāgarī and other scripts):  $\mathbf{..} \sqcup (h + p)$  for [f] and  $\mathbf{..} \mathbf{20}$  (h + j) for [z].

# 9.2.7. Sinhalese

Writing was introduced to Sri Lanka by the 2nd century BCE (and possibly as early as the fourth). The subsequent development of the Sinhalese script was characterized by long periods of isolation, interrupted by occasional strong influence from mainland South Indian scripts (especially Pallava Grantha), and by the 14th century it had approached its modern form. It has special signs for the Sinhalese open vowels  $\ddot{a}$  ( $\dot{q}_{1}, -\tau$ ) and  $\ddot{a}$  ( $\dot{q}_{1}, -\tau$ ) and the prenasalized stops  $\check{n}ga$  ( $\infty$ ),  $\check{n}da$  ( $\hat{\omega}$ ),  $\check{n}da$  ( $\hat{q}$ ), and  $\check{m}ba$  ( $\hat{w}$ ). The Sinhalese character repertoire can be divided into two layers: the core characters for Classical Sinhalese (elu hōḍiya) and an outer set with signs for Sanskrit and Pāli words (r,  $\bar{r}$ , ai, au, the aspirates, the nasals  $\dot{n}a$  and  $\tilde{n}a$ , and the sibilants  $\dot{s}a$  and sa); the complete alphabet is called miśra hōḍiya. Letters added in the miśra hōḍiya are in normal speech pronounced like corresponding letters from the elu hōḍiya ( $\varpi$  gha and  $\omega$  ga are both [gə], etc., and the sibilants are all [s]).

# 9.2.8. Dhivehi

Dhivehi, spoken on the Maldives, is closely related to Sinhalese, and was first committed to writing in the 13th century in the evēlā ('ancient') script derived from Sinhalese, developing further into the *dhivehi akuru* ('island letters') script (Geiger 1919: 20-29, 149-168, DeSilva 1969). In the early 17th century this was replaced by the current script, called Thaana, which is written from right to left and contains 24 consonant letters. The first nine are based on the Arabic numerals; the next nine on an old set of local numerals; and the last six letters, used for loanwords, are modifications of other letters or borrowings from Arabic. There are ten vowel signs, including one for short a (-) which is not inherent in consonant signs. The absence of a vowel is redundantly signalled by a cancellation mark ( $\stackrel{\bullet}{}$ , called sukun). The character alifu ( $\land$  ') serves as vowel carrier for initial and post-vocalic vowels. Alifu, r š, or r in combination with sukun ( $\mathring{a}$ ,  $\mathring{r}$ ,  $\mathring{r}$ ) represent a glottal stop word-finally; when preceding another consonant, they indicate gemination, and مُرْضُ also adds an off glide [j] to the preceding vowel: مُؤَمَّضُ ba'te' [batte?] 'eggplant', مُوَثِرُ raš [ra?] 'island', مُوثِرُ atpulu [ajppulu] 'hand'; prenasalization is marked by  $\sim n$  (without sukun) and sometimes left unmarked:  $\dot{z}(n)$ ,  $\dot{v}$  ka( $\check{n}$ ), du (examples from Gair & Cain 1996). Diphthongs are written as vowel + alifu + vowel (e.g. زم fa'i [fai] 'leg'). Loanwords from Arabic are written in Arabic script or with the help of twelve additional characters formed by adding dots to Thaana letters.

# 9.2.9. Latin script

The only South Asian language using (since the 16th century) the Latin script as its primary writing system is Konkani, where vowel length is not marked and retroflexes are indicated by double consonants. Attempts to replace other South Asian scripts by the Latin script were unsuccessful, reflecting strong attachment to and identification with the local scripts as well as the effectiveness of the Brāhmīderived scripts in representing Indian languages.

# 9.2.10. Numerals and punctuation

Early Kharoṣṭhī had number signs for 1, 10, 20, 100, and 1000 (1, 2, 3, 4, 2); 20 is a cursive combination of two signs for 10 arranged on top of each other. Later a separate sign for 4 (x) is added to the inventory and cursive forms for 2 and 3 develop (P from 11, P from 111). The Kharoṣṭhī number system is additive, with higher number signs preceding lower ones in the reading direction: 16, e.g. is written 10 (+) 4 (+) 1 (+) 1 (11x2). Multiples of 100 and 1000 are written with a multiplier preceding the 100 or 1000: 200 is 411. The Kharoṣṭhī system is based on the Aramaic one (Chrisomalis 2010: 68–74, 83–86).

The early Brāhmī number system is also additive, but has a larger number of basic signs for 1 to 9 ( $\neg$ , =, =, ?, ?, ?, ?, s, s), for 10 to 90 ( $\propto$ ,  $\bullet$ , v, H, z, s, s,  $\bullet$ ) and for 100 and 1000 ( $\neg$ ,  $\bullet$ ). Higher signs precede lower ones in the reading direction: 16 is 10 (+) 6 ( $\propto r$ ). Multiples of 100 and 1000 are written with a multiplier following and conjoined with the 100 and 1000 (r, r is 1004, but r is 4000); 200/2000 and 300/3000 are written by adding one or two horizontal strokes ( $\neg$ , r, and  $\neg$ , r). The origin of the early Brāhmī number system remains unclear, but inspiration from China (Falk 1993: 168–176) or Egypt (Chrisomalis 2010: 191–192) may have played a role. Around the 7th century, the positional system came into use, with the signs for 1 to 9 continuing those of the older system and a new sign for 0 (a dot or circle).

Some of Aśoka's edicts use word spacing to mark syntactic units (Janert 1972). Other inscriptions and manuscripts are written continuously, but use various punctuation marks such as ',  $\circ$ ,  $\circledast$  and  $\in$  in Kharoṣṭhī and  $\circ$  and  $\circ$  in Brāhmī, and later the signs daṇḍa (I) and double daṇḍa (II). Since the early 20th century, European punctuation has increasingly been used in Indian texts.

# **9.3. Recent script-related research** By *Stefan Baums*

# 9.3.1. Recent work on alphasyllabic writing systems

The two ancient writing systems of South Asia, Kharoṣṭhī and Brāhmī, were first deciphered with the help of bilingual coin legends and by working forwards from the letter shapes of the Semitic scripts (Kharoṣṭhī) and backwards from those of later South Asian scripts (Brāhmī). After pioneering efforts by Charles Wilkins (1749–1836), Henry Thomas Colebrooke (1765–1837), Christian Lassen (1800–1876), and others, James Prinsep (1799–1840) announced his decipherment success in two articles — in 1837 on the inscriptions of Sanchi, and 1838 on the Indo-Greek coin legends. The decipherment of Kharoṣṭhī was consolidated when Norris (1846) published his reading of the newly-discovered Aśokan edict at Shabazgarhi (Falk 1993: 99–103; Salomon 1998: 199–215). From the midnineteenth until the beginning of the twentieth century, knowledge of the historical scripts of South Asia solidified and received first synthetic treatments by Dowson (1863) on Kharoṣṭhī, Burnell (1874, 1878) on the South Indian scripts, and Bühler (1896) in his comprehensive paleography.

The modern study of the historical scripts can be divided into three formative phases. The first of these was prompted by the discovery of large numbers of early Buddhist manuscripts from South Asia along the Silk Roads in modern Xinjiang, China, and near Gilgit in modern Pakistan. Hoernle (1916) and Boyer, Rapson, Senart & Noble (1920–1929) provided the first major publications and analyses of the Central Asian Brāhmī and Kharoṣṭhī material; v. Hinüber (1979) summarizes research on the Gilgit manuscripts. Next, the mid-twentieth century saw a number of new syntheses by Das Gupta (1958) on Kharoṣṭhī, Dani (1963, 1986) in a new comprehensive paleography giving particular attention to regional developments of Brāhmī, Sircar (1942, 1965a, 1965b, 1966, 1983) in a series of reference works on South Asian epigraphy, Sander (1968) on the development of Brāhmī in Central Asia, and Jensen's (1969) overview of the South Asian scripts as part of a general history of writing systems.

Finally, another series of compendia inaugurated the third and current phase of historical studies: v. Hinüber (1990) and Falk (1993) reevaluated our knowledge of writing and literacy in ancient South Asia and assembled a comprehensive history of research (cf. the review article by Salomon 1995). Daniels & Bright 1996 gave a new overview of the world's writing systems, replacing Jensen's book with a collection of essays. Salomon's (1998) handbook of South Asian epigraphy similarly updated Sircar 1965a, and Falk & Slaje (eds.) 2000–2005 broke new ground with a paleographical database assembling the contributions of numerous experts; Einicke (2009) draws on this database for a comprehensive handbook of South Asian scribal notation from the fifth century to modern times. The third phase,

like the first, is also characterized by the emergence of large amounts of new primary material. Recent discoveries started off with Coningham, Allchin, Batt & Lucy's (1996) report of a find of potsherds from Anurādhapura, Sri Lanka, dated stratigraphically to the early fourth century BCE and thus potentially indicating a history of Brāhmī as a traders' script before its adoption by Asoka. The discovery of around eighty-five Gāndhārī manuscripts on birch bark and palm leaf (Salomon 1999; Strauch 2008; Baums & Glass ongoing) has put Kharosthī manuscript studies on an entirely new footing (Glass 2000 discusses the paleography of this material, Baums 2009: 110-200 its orthography). The recovery of large numbers of first-millennium Sanskrit manuscript fragments from Gilgit and Bamiyan (Hartmann 2000, Braarvig 2000) and of early second-millennium Sanskrit manuscripts from Tibet (Steinkellner 2004) is filling gaps in our knowledge of the development of Brahmi (Sander 2000) and the Transitional Scripts. The gradual replacement of the Kharosthī by the Brāhmī tradition and the concomitant switch from Middle Indo-Aryan to Sanskrit for the transmission of Buddhism is discussed by Salomon (2008) and Strauch (2012). The decipherment of the rare Bhaiksukī (or Arrow-Headed) script of the Transitional period has been completed with the help of a manuscript discovered in Nepal (Dimitrov 2010). The Shell Script remains one of the undeciphered writing systems of South Asia (Salomon 1987).

With the rise of writing-system studies as a part of modern linguistics, new analyses and questions were brought to bear on the South Asian writing systems. The traditional typology of writing systems as either alphabetic, syllabic, or logo-graphic could not accommodate Kharoṣṭhī and the Brāhmī-derived scripts, and a fourth type — called "alphasyllabic" or "abugida" — was defined for the purpose (Bright 1994: 323–324; Bright 1999; Salomon 2000; Coulmas 2003: 131–150; Swank 2008). Building on the formal analysis of Sproat (2006a), Weingarten 2011 suggests that the typological category of the South Asian scripts varies by specific feature examined and perspective (semasiological or onomasiological) adopted.

# 9.3.2. Script and literacy

Modern studies of script acquisition and literacy in South Asia can be traced back to development studies and the "comparative reading" approach of the 1970s and 1980s (Oommen 1973; Malmquist 1982; Hladczuk & Eller 1987). Examples of recent work are Karanth & Suchitra 1993 on Hindi and Kannada, Patel 1995 on Gujarati, Prakash & Joshi 1995 on Kannada, and the 2004 special issue of the journal *Reading and Writing* on 'reading and writing in semi-syllabic [= alphasyllabic] scripts' with contributions like Vasanta 2004 (Telugu), Gupta 2004 (Hindi), Karanth, Mathew & Kurien 2004 (Kannada), and Chengappa, Bhat & Padakannaya 2004 (Hindi and Kannada). The question of orality and literacy and their relative spheres of application in South Asia is addressed by Bright (1990: 130–146);

Patel (1993) (on ancient South Asia; cf. v. Hinüber 1990); Glück (1994: 741–742) (on the relationship of literacy and diglossia); Jain (2003: 50–53); and Agnihotri (2008). The effects of South Asian "multigraphism", "digraphia", or "bi-literacy", i.e. the use of more than one script by an individual, are studied by Ferguson (1978); Pederson (2003) (Tamil-English biliterate readers have more precise shape recognition than monoliterate readers); Prakash et al. (1993) (Kannada-English and Hindi-English readers perform better at phonemic segmentation); Wali et al. (2009); Unseth (2005: 36–37); Vaid (1995) (script directionality influences production and perception of non-linguistic shapes).

# 9.3.3. South Asian scripts in Unicode

After birch bark, palm leaf, and paper, the South Asian writing systems are now undergoing a further transition to digital encoding as one of their primary media. The Unicode Consortium (representing industry, academic, and governmental interests) is responsible for defining the universal digital coding system of the world's scripts (Unicode Consortium 1991-; Baums 2006: 111-116), and the first version of the Unicode standard (1.0, released in October 1991) provided support (modelled on the 1988 ISCII standard) for the nine major modern indigenous scripts of India: Devanagari, Bengali, Gujarati, Gurmukhi, Kannada, Malayalam, Oriya, Tamil, and Telugu. Eight years later (3.0, September 1999) the standard added support for Sinhala and Thaana, completing coverage of the major modern indigenous scripts of South Asia. Further South Asian scripts were added as follows: in March 2005 (4.1), Kharosthī and Syloti Nagri (a nineteenth-century modification of the Bengali script for the Syloti dialect); in April 2008 (5.1), Saurashtra (developed in the nineteenth century on the basis of Gujarati and Oriya, and used by Gujarati immigrants to southern India) and Ol Chiki (an alphabet developed in 1925 by Raghunath Murmu for the Munda language Santali; Zide 1996: 612); in October 2009 (5.2), Kaithi (a cursive form of Nagari used by traders from the 16th to the early 20th century) and Meetei Mayek (developed for the Tibeto-Burman language Manipuri and used until the early 18th century); in October 2010 (6.0), Brāhmī; and in January 2012 (6.1), Śāradā, Takri, and Sorang Sompeng (an alphabet invented in 1936 by Mangei Gomango for the Munda language Sora; Zide 1996: 613). The Unicode standard thus currently contains support for six historical scripts, the eleven major modern indigenous scripts, and four modern minority scripts. The most urgent desiderata of the standard are coverage and usage guidelines for the remaining historical scripts, informed by a proper historical classification and meeting the practical needs of the scholarly community, and the implementation of software support for all covered scripts.

# 9.3.4. The Indus inscriptions

The inscriptions of the "Indus" or "Harappan" civilization (Wheeler 1968; Kenoyer 1998) on seals and other objects have long been considered the earliest — and as yet undeciphered — writing system of South Asia (Possehl 1996 provides a good overview). A precursor to the sign system of the seal inscriptions were potters' marks, used by the people of the early Indus civilization from the fourth millennium BCE to about 2600 BCE. Then, after a relatively short transition period, the fully developed Indus sign system came into being by 2500 BCE. It disappeared with the decline of the Indus civilization around 1900 BCE (1700 BCE in its southern outposts in Maharashtra).

The overall inspiration for the development of the Indus sign system may have come from the Indus people's western trade partners in Mesopotamia (trade relations are attested by around 40 Indus seals found in the Near East). The Indus civilization did not, however, borrow the Mesopotamian cuneiform, but invented their own sign system. Some of the shapes of the Indus signs appear to point back to the earlier potters' marks.

More than 4,000 inscriptions in the Indus sign system are known today, most of them seal inscriptions, some on amulet tablets and pottery. The corpus of known Indus inscriptions is catalogued by Joshi & Parpola (1987), Shah & Parpola (1991), and Parpola, Pande & Koskikallio (2010), building on earlier work by Mahadevan (1977).

It is reasonable to suppose that the referents of the seal inscriptions are similar to those of Mesopotamian seals: items of merchandise and the names of owners and titles of office, often incorporating the names of gods. The average length of the inscriptions is just five signs, ranging from single-sign inscriptions to an untypically long 28-sign inscription on the sides of a prismatic amulet. The total number of different signs in these texts is roughly 400. If the Indus sign system does in fact represent a full writing system (see below), then this relatively high number together with the pictorial nature of most signs would point to a logographic writing system. That the number of signs is not even larger may have to do with the specialized nature of the texts.

Since the first publication of an Indus seal in 1875 (Cunningham 1875: 108, pl. XXXIII), many attempts have been made to decipher the Indus inscriptions as a writing system, but none of them is fully convincing. Parpola (1994, 1996) summarizes earlier attempts and discusses the challenges to interpreting the Indus inscriptions. Our understanding is hampered by the brevity of the texts, the absence of parallel texts in another writing system, and our ignorance of the language (if any) that is used in the Indus inscriptions. Beyond that, many proposed decipherments also suffer from methodological weakness. Equating Indus signs with similar-looking characters in other ancient writing systems is, for instance, a doubtful procedure due to the arbitrariness of similarity judgments and the fact

that different writing systems tend to use the same basic geometric shapes for different purposes. More promising are attempts starting from distributional criteria. Using these, it has been argued that a certain class of signs probably represents suffixes or phonetic/semantic determinatives. Another class of characters, consisting of groupings of vertical lines, probably represents the numerals of the language iconically. These numeral signs typically precede a limited class of other signs, probably denoting the things being counted; if the inscriptions do write a language, then the position of the numerals in front of their headword would give a typological clue to the type of language represented. (It seems relatively certain, from the stroke order of characters and their spacing in lines, that the Indus inscriptions were incised from right to left.)

The strongest contender for a language underlying the Indus inscriptions is Dravidian, as suggested in the work of Yurii Knozorov (e.g. 1965), Asko Parpola (e.g. 1994, 1996, 2008), and Iravatham Mahadevan (e.g. 1977, 2003). There are pockets of speakers of Dravidian in Northern India and Baluchistan, indicating an early area of use much larger in ancient times than today; this is confirmed by the presence of Dravidian loan-words in the Rgveda, composed in the Indus region in the second millennium BCE.<sup>3</sup> Some of the distributional patterns of Indus signs are also reminiscent of patterns of homophony in Dravidian languages. It is a priori less likely that the Indus inscriptions record an early Indo-Aryan dialect (recent suggestions to this effect include Rao 1982 and Jha & Rajaram 2000), since we have no evidence that Indo-Aryan was used in South Asia when the inscriptions were produced.

A radically different approach to the interpretation of the Indus inscriptions has recently been introduced by Farmer, Sproat, and Witzel (2004), who argue that the Indus inscriptions reflect a non-linguistic symbolic system. They adduce the inscriptions of the southeast European Vincha complex (38-39) and the Near Eastern emblems used, e.g. on boundary stones (39-40), as ancient parallels for such non-linguistic symbolic systems, adding the medieval Scottish heraldic system as a later example (27-28). In their interpretation, the function of the Indus inscription was not the conveying of linguistic messages, but the association of important natural, supernatural, and social entities (40-43) in a cohesive ideological system operating in a dispersed multilingual population (45). Contrary to traditional interpretations of the Indus inscriptions, even number signs may not have served accounting purposes, and may sometimes have been used metaphorically (e.g. to refer to numeric sets of deities; 41-42). From the archeological absence of longer inscriptions on nonperishable material and of writing utensils, Farmer, Sproat, and Witzel argue further not only that the known Indus inscriptions are non-linguistic, but that the Indus civilization had no writing system at all.

<sup>&</sup>lt;sup>3</sup> See also Section 2.3 for a different view, and 1.6.1.2 for recent Dravidologist views on Brahui.

In a rejoinder to Farmer, Sproat, and Witzel's suggestion, Parpola (2008) continues to maintain a linguistic (and most likely Dravidian) interpretation of the Indus inscriptions. He particularly stresses that the alternation of different number signs before an identical non-number sign does point to counting (116); that cotton cloth was one of the main trade goods of the Indus civilization, and according to Nearchus was used for writing in the 4th century BCE, but that no ancient specimens of cloth are preserved from the Indus area, making it less unlikely that longer Indus texts on perishable material existed (117); that brushes were evidently used for inscribing pots and possibly also for manuscripts (118-119); and that seal impressions have been found on clay tags that were probably attached to merchandise, indicating a commercial use of some Indus seals after all (122). The work of Yadav and Vahia (2011) illustrates the ongoing formal analysis of the Indus inscriptions on the assumption that they do represent a writing system. While no scholarly consensus on the status of the inscriptions as writing or non-linguistic symbols and (if the former) on the language of the Indus civilization has yet emerged, the arguments have reached a level of refinement and methodological reflection that will hopefully result in a clearer definition of the possible scope and limits of our knowledge.

# 9.4. Perso-Arabic adaptations for South Asian languages By Elena Bashir

9.4.1. Early adaptations

The Arabic script consists of 28 consonant letters, three of which can indicate both consonants and long vowels, short vowels being (optionally) indicated with diacritics. Like scripts derived from it, it is cursive and has no lower-/upper-case distinction. When the Arabic script was adopted for other, non-Semitic languages, various kinds of modifications became necessary.<sup>4</sup> For writing Persian, the 28 original Arabic characters were supplemented by the addition of  $i, \mathfrak{F}, \mathfrak{G}, \mathfrak{G}, \mathfrak{and} \downarrow, [\check{z}], [\check{c}], [g], and [p], respectively. Later, Urdu required representing the phonological distinctions between retroflex and dental and between aspirated and unaspirated consonants, and a unique final <math>/\tilde{e}/( \)$  to indicate grammatical distinctions. Parvez (1996: 15) notes that these early extensions of the Arabic script exhibited partial systematicity, i.e. three dots below to indicate voiceless sounds, e.g.  $\downarrow$  [p],  $\mathfrak{F}$ [č]. Not all voiceless sounds, however, have three dots below, e.g.  $\neg$  [t]. In Urdu, the representation of retroflexion by a small  $\flat$  diacritic above, and of aspiration by a digraph consisting of stop consonant +  $\omega$ , is consistent.

<sup>&</sup>lt;sup>4</sup> See 9.2.8 above for discussion of the Arabic-based script used for Dhivehi.

Panjabi has been written in the Perso-Arabic script ("Shahmukhi") since the 12<sup>th</sup> century, using the same set of characters as Urdu. Panjabi, however, is a tone language, while Urdu is not; original voiced aspirates have changed in Panjabi to either voiceless unaspirated or voiced unaspirated stops, depending on their position. Syllable-initial Urdu voiced aspirated stop letters represent Panjabi voice-less stops with low tone on the following vowel, e.g.  $\forall \downarrow p ab\bar{l}$  'brother's wife'. Word-final Shahmukhi voiced aspirate letters indicate unaspirated voiced stops with high tone on the preceding vowel, e.g.  $\forall \downarrow p ab\bar{l}$  'holy man, ascetic'. The two other letters representing consonant /h/,  $\bullet$  and  $\zeta$ , also indicate tone in Panjabi, e.g.  $\forall \downarrow p a \bar{l}$  'tenth month of Bikrami calendar', or  $\zeta \downarrow [nik a]$  'Muslim marriage contract'. Urdu also lacks /n/ and /l/. Panjabi's phonemic /n/ and /l/ are still not uniquely represented in Perso-Arabic used for Panjabi. Various proposals for representing them in Unicode are under discussion (Malik 2005).

Kashmiri has been written in the Arabic script since the 15<sup>th</sup> century. The consonants are the same as those used for Urdu, but the vowel symbols have been considerably augmented by various diacritics to represent Kashmiri vowels not found in Urdu (Koul 1995). The orthography is still not standardized.

The earliest known Pashto manuscript is dated to the mid-17<sup>th</sup> century, but it is not known when the current standard orthography was adopted (MacKenzie 1997). In the current standard, retroflexion is marked with a ring attached to the body of the letter, e.g.  $\downarrow$ /d/, and there are two "versatile" consonants,  $\downarrow$  and  $\omega$  for voiced and voiceless fricatives respectively, which are pronounced [z], [ $\check{\gamma}$ ], or [ $\gamma$ ] (voiced retroflex fricative, voiced palato-velar fricative, or voiced velar fricative); and [ $\S$ ], [ $\check{x}$ ], or [x] (voiceless retroflex fricative, voiceless palato-velar fricative, or voiceless velar fricative), respectively, according to a speaker's dialect (MacKenzie 1959: 232). Some orthographic differences exist between Afghan and Pakistani Pashto; for Afghan Pashto see Penzl 1954. Mirdehghan 2010 compares the representations of consonantal and vocalic sounds in Persian, Urdu, and Pashto and their orthographic systems.

The present Sindhi script, including 52 characters and seven diacritics, was instituted by the British in the 1850s. It represents aspiration and retroflexion inconsistently, e.g.  $_{2}$  [dh] but  $_{u}$  [bh], and  $_{2}$  [d], but  $_{u}$  [t]; three of the implosive consonants are represented consistently with two vertical dots —  $_{u}$  [6],  $_{\overline{c}}$  [ $_{1}$ ], and  $_{\zeta}$  [d] — but the fourth,  $_{2}$  [d], is not. Khubchandani (2003: 635) gives a complete list of Sindhi letters. All the Sindhi characters in use have Unicode code points.<sup>5</sup>

Though Balochi has been written since at least 1873 (Jahani 1989: 23), and though questions of orthography and spelling have been hotly debated for years, 'no one orthography has won general acceptance among the Balochi cultural elite' (Jahani & Korn 2009: 638). Jahani (1989) discusses the historical, political, and

<sup>&</sup>lt;sup>5</sup> For an early discussion of the introduction of Unicode for languages of Pakistan, especially Sindhi, see Bhurgri MS.

linguistic issues in great detail. Unique developments in Balochi orthography are morphophonemic symbols, for instance the use of ; for the oblique singular case ending /ā/ and , for the genitive suffix /əy~ī/ (Barker & Mengal 1969, Vol 2: 9, 37–39). An attempt was made in the early 1990s to introduce four new hybrid "cross-dialectal" symbols that could represent dialectal variation between Eastern and Western Balochi, respectively —  $\oint [f/p(h)]$ ,  $\oint [\gamma/g]$ ,  $\int [x/k]$ , and  $\oint [\theta/t]$ — which would function like the Pashto "versatile" letters  $\psi$  and  $\int (Balochistan$ Textbook Board 1989, Barker & Mengal 1969, Vol 2: 8). Speakers of Eastern dialects would pronounce the first (fricative) variant, while speakers of other dialectswould say the second (stop) variant. This interesting proposal, however, has notsurvived.

Brahui began to be written in the 19<sup>th</sup> century, but literary production received impetus after Independence in Quetta (Elfenbein 1983: 107). Now since the 1960s it is written with the same Urdu-style symbol set as Balochi, but with the addition of  $\hat{J}$  (U+06B7) to represent the voiceless lateral fricative [1] (Balochistan Textbook Board 1991). In 2008 a Brahui Language Board was established; one of its tasks was to be redesigning Brahui script for Brahui. One of their efforts can be seen at https://sites.google.com/site/brahuilb/ (this orthography is different from that in Balochistan Textbook Board 1991). Now fonts and keyboard layouts have been developed for Brahui, facilitating its use in the modern world.

Khowar has been written, using Persian (later Urdu) orthography, since the 17<sup>th</sup> century, first in a mixture of Persian and Khowar (Bashir 2006). In the early 20<sup>th</sup> century, symbols were devised by Prince Hussam ul-Mulk and his son Samsam ul-Mulk (ul-Mulk, n.d.) for the Khowar consonant sounds not found in Persian or Urdu: /s/, /c/, /j/, and /z/. These symbols have remained in use, and are now encoded in the Unicode Standard as (U+0770), (U+076F), (U+076E), and (U+0771). Letters for /ts/ and /dz/ were already encoded as (U+0685) and z' (U+0681), respectively (www.unicode.org/charts). Buddruss (1982) analyzes the considerable orthographic variation obtaining at the time of his writing; it seems that with increasing writing and publication in Khowar the orthography is slowly moving in the direction of standardization.

# 9.4.2. Recent adaptations

Several recent adaptations reflect modern linguistic analysis. These include those for Torwali, Kalasha, Burushaski, Wakhi, Shina, Gawri, and Pashai. Many northwestern languages have phonemic tone (Baart 2003), which is only sometimes represented. Panjabi (though indirectly), Burushaski, and Kohistani Shina (Schmidt & Kohistani 1995) represent tone in their writing systems, using different techniques. Debate on whether and how to represent tone in other varieties of Shina and in Khowar continues. The advent of Unicode, coinciding with increased concern about language endangerment and language documentation, has given a new impetus to work on developing scripts for previously unwritten languages. Recently new Unicode characters were proposed for Khowar, Torwali, and Burushaski (Bashir, Hussain & Anderson 2006) and have been added to the Arabic Supplement code page.

# 9.4.2.1. Shina

The first published attempt at accurately representing Shina's phonology was by Namus (1961: 28–29), who introduced eight new consonant symbols and a system of vowel diacritics indicating four degrees of length: slight, short, normal, and long. Zia (1986, 2010) and Taj (1989) have since followed, each with a different proposal. Since the beginning, debates on (Gilgit) Shina orthography have continued to focus on the questions of whether or not it is necessary to represent vowel length and phonemic tone. Schmidt & Kohistani (1995 ms: 5) discuss the analysis underlying their work on an orthography for Kohistani Shina, focusing on the question of representing tone. They conclude that, 'It is possible to use length to predict stress and the occurrence of tone, and this approach is more appropriate to the consonant-rich Arabic orthography.' The article includes a list of symbols used in their scheme.<sup>6</sup> Buddruss 1983 is a history of the development of writing in Shina.

# 9.4.2.2. Burushaski

Burushaski began to be written in the 20<sup>th</sup> century through the efforts of Allama Nasir ud-Din Nasir Hunzai. Some early publications employed roman representations (e.g. Hunzai n.d.); recently, however, a Perso-Arabic representation is being employed by the Burushaski Research Academy in their dictionary project (Burushaski Research Academy 2006, 2009, 2014). A list of characters used and their phonetic values can be seen in Burushaski Research Academy 2011. The vowel symbols do represent tone.

# 9.4.2.3. Saraiki

Development of a specifically Saraiki writing system has begun within the last quarter century. Both Saraiki and Sindhi have four voiced implosives: [6], [f], [d], and [g]; usually represented as  $\downarrow$ ,  $\sharp$ ,  $, \sharp$ , respectively, although these representations are not yet entirely standardized. The existing Sindhi characters for

<sup>&</sup>lt;sup>6</sup> This paper was subsequently published in Israr-ud-Din (ed.) 2008, *Proceedings of the Third International Hindu Kush Cultural Conference*, 283–287, Karachi: Oxford University Press. However, the published version introduces errors not present in the original 1995 manuscript. The reader is advised to consult the original, at http://www.hf.uio.no/ikos/english/research/projects/shina/publications/Schmidt%20and%20 Kohistani%202008-original-1.pdf.

these sounds are not employed in the same way. Several suggestions for other representations are current, but none has won acceptance. Some of these can be seen in Shackle 2003: 598. The Unicode character U+0768  $\overset{*}{_{U}}$ , has been included for Saraiki and Potohari [n]. Rasoolpuri 1976 is a short history of Saraiki orthography.

# 9.4.2.4. Kalasha

Several different schemes for writing Kalasha have been put forward — some employing roman script and others Perso-Arabic. Trail & Cooper 1999 is a dictionary using a system devised by its authors in collaboration with the Kalasha community. Following the Urdu convention for representing retroflex consonants, it represents retroflex vowels also with a small diacritic  $\downarrow$  above the vowel symbol. Several unique consonant symbols have also been devised, but to my knowledge, these have not yet been incorporated in Unicode. Heegård 2000 discusses the interplay of political and linguistic factors in the designing of alphabets for this language.

# 9.4.2.5. Torwali

Inam Ullah (2004) reports on his work on developing a writing system for Torwali. His system reflects phonological analysis and distinguishes the retroflex sibilants and affricates and the low front vowel i [æ] (see Inam Ullah n.d.). His *Torwali-Urdu dictionary* (Inam Ullah 2010) and online Torwali dictionary (http://www.cle. net.pk/otd/) employ this system.

# 9.4.2.6. Gojri

Losey 2002 is a phonological analysis of Gojri, done to provide a foundation for script development efforts; it includes discussion of the representation of tone and a list of the characters used and their phonemic values.

# 9.4.2.7. Gawri

Work on developing a script and orthography for this language, spoken in the upper reaches of the Swat Valley, has made rapid progress since 1995, when a Kalam-based Spelling Committee was established. Linguistic considerations and conventions for indicating consonant sounds not present in Urdu, tone, and vowel length are discussed in Baart & Sagar n.d.: 8–10. Sagar 2008 discusses literacy efforts and publications using the Perso-Arabic script variant adopted. The Gawri characters can also be viewed in Inam Ullah n.d.

# 9.4.2.8. Pashai

Pashai has until very recently been unwritten. Current efforts to create a standard orthography draw from both the locally-organized Darrai Nur Language Committee (DNLC) and the Minority Language Committee at the Afghan Ministry of Education. Both suggested orthographies are based on the Pashto script (Perso-Arabic), and use the same notation for retroflex consonants. The Darrai Nur Language Committee's orthography is based on a phonemic analysis of Pashai, and is intended for adult literacy training, as opposed to the Ministry's version, which retains historical spellings for borrowed words. A new character adopted for Pashai is j (U+06B5) to represent [4] (Rachel Lehr p.c. 3 Dec. 2014; see also Yun 2003). Lamuwal & Baker 2013 contains two sample texts of a well-known folk tale — one in the DNLC orthography and the other in the orthography developed by the Ministry of Education.

# 9.4.2.9. Wakhi

Roman/IPA (Ali 1980) and Perso-Arabic (Sakhi 2000) script alternatives have been advocated for Wakhi in Pakistan. In Pakistan, the roman/IPA approach seems to be predominating, while in Afghanistan Perso-Arabic Pashto orthography has been adopted, and in Tajikistan Cyrillic is used (Mock 1998: 36–37). Beg, Mock & Wakhani 2014 is a detailed discussion of recent developments in orthography debates, and of computer fonts and keyboards for Wakhi.

# 9.4.2.10. Balti

Starting in 727 CE, when Baltistan was conquered by the Tibetans, writings in Balti, a TB language, were in the Tibetan script; in the 16<sup>th</sup> century the Perso-Arabic script was introduced (https://baltistaan.wordpress.com/category/history/). Several orthographies have been employed for (contemporary) Balti. Sprigg (1996) describes his development of a roman-based orthography and dictionary for Balti, the phonological representations in which are intended to facilitate comparison with Classical Tibetan. This effort culminated in Sprigg's (2002) dictionary. Recently there has been a local initiative by literary scholars and social activists working through the Baltistan Cultural Foundation to revive the Tibetan script for Balti, in an attempt to preserve indigenous Balti culture and ethnic identity.<sup>7</sup> See

On the request of local activists, the September 2006 meeting of ISO/IEC 10646 WG2 agreed to encode two new characters in the Tibetan block — 0F6B Tibetan letter KKA, and 0F6C Tibetan letter RRA — in order to facilitate writing Urdu loanwords used in modern Balti using Tibetan script (https://baltistaan.wordpress.com/category/history/).

Kazmi 1996 and Khan 2000 for discussion of these issues. The Baltistan Cultural Foundation has published a Tibetan-script primer for Balti, and has encouraged the use of this script on local signboards.<sup>8</sup> As of 2002, however, the Perso-Arabic alternative seemed to have retained its dominant position. National Language Authority 2002 is a Perso-Arabic Balti primer, which includes five non-Urdu characters. Chitrali 2004 is another, privately-published, Perso-Arabic primer, which includes another six non-Urdu characters.

# 9.4.3. Diverse representations

Several languages have phonetically similar/identical sounds, but their speakers, presumably wanting to maintain cultural uniqueness, have chosen to use separate characters for them. For example, the retroflex voiceless sibilant [s] is found in several languages of northwestern Pakistan, but it is represented differently in each of them: Kohistani Shina uses the basic shape for  $\tau$  with two short horizontal lines above it (no Unicode code yet); Khowar uses  $\dot{\tau}$  (U+0770); Burushaski uses (U+077D); Torwali  $\ddot{\tau}$  (U+075C); Kalasha uses  $\dot{\tau}$  with a small  $\flat$  diacritic above it (no Unicode code yet); and Gowri uses).

# 9.5. New research areas and desiderata By *Elena Bashir*

Currently research on writing systems is very active. According to Sproat (2000: 127), 'The question of what kinds of linguistic elements written symbols represent is the single most investigated issue in the study of writing systems.'<sup>9</sup> Current research moves past earlier classifications like the "deep : shallow" distinction and asks other, new questions.<sup>10</sup> Veldhuis & Kurvers 2012 raises the question of how the acquisition of writing affects language processing (in the brain); and Banga et al. 2012 asks whether knowledge of the relationship between speech and writing in one language influences understanding of that relationship in another. This question is particularly relevant for South Asia, where common scripts are shared by languages with many layers of historical accretion, convergence, and divergence.

<sup>&</sup>lt;sup>8</sup> Pandey 2010 is a proposal to the Unicode Consortium recommending yet another script for writing Balti.

<sup>&</sup>lt;sup>9</sup> A conference, Signs of Writing: The Cultural, Social and Linguistic Contexts of the World's First Writing Systems, was held at the University of Chicago on 8–9 November 2014.

<sup>&</sup>lt;sup>10</sup> Sproat 2000: 128–144 contains an overview of various taxonomies of writing systems and many references.

#### 810 Bibliographical References

Orthographic conventions like spelling, even when generally agreed upon, are often not consistently applied. Spelling and punctuation variation are fertile fields for study, especially as computational approaches to language processing and analysis make them more feasible. Other emerging fields of research are the interesting use of roman to represent South Asian languages in email and text messaging, and the representation of code mixing and code switching in writing; see Sebba, Manootian & Jonsson (eds.) 2012. What may such roman and/or mixed representations reveal about users'/writers'/speakers' understandings of their languages and of themselves? See Section 9.3 above for discussion of three other important areas of current research.

# **Bibliographical References**

Agnihotri, Rama Kant	
2008	Orality and literacy. In: Braj B. Kachru, Yamuna Kachru, and S. N. Sridhar (eds.),
	Language in South Asia, 271–284. Cambridge: Cambridge University Press.
Ali, Haqiqat	
1980	xəkwör zik [Wakhi language], Book 1. Gilgit: Wakhi Culture Association.
Asani, Ali	4
1987	The Khojkī script: A legacy of Ismaili Islam in the Indo-Pakistan subcontinent.
	Journal of the American Oriental Society 107: 439–449.
Baart, Joan L. G.	
2003	Tonal features in the languages of northern Pakistan. In: Joan L. G. Baart and
	Ghulam Hyder Sindhi (eds.), Pakistani languages and society: Problems and
	prospects, 132-144. Islamabad: National Institute of Pakistan Studies, Quaid-
	i-Azam University/Summer Institute of Linguistics.
Baart, Joan L. G., and Muhammad Zaman Sagar	
n.d.	The Gawri language of Kalam and Dir Kohistan. http://www.fli-online.org/
	documents/languages/gawri/gawri_introduction.pdf (accessed 5 December
	2014)
Balochistan Textbook Board	
1989	Baločī qāeda [Balochi primer]. Quetta: Qalat Publishers.
Balochistan Textbook Board	
1991	Brāhūī qāeda [Brahui primer]. Quetta: Mahmud Stationers and Book Sellers.
Banga, Arina, Esther Hanssen, Robert Schreuder, and Anneke Neijt	
2012	How subtle differences in orthography influence conceptual interpretation.
	Written Language & Literacy 15(2): 185–208.
Barker, Muhammad Abd-al-Rahman, and Aqil Khan Mengal	
1969	A course in Baluchi, volumes I and II. Montreal: Institute of Islamic Studies,
	McGill University.
Bashir, Elena	
2006	Indo-Iranian frontier languages. In: Iranica [online] http://iranica.com/arti-

cles/indo-iranian-frontier-languages-and-the-influence-of-persian (accessed 5 December 2014) Bashir, Elena, Sarmad Hussain, and Deborah Anderson

2006 Proposal for characters for Khowar, Torwali and Burushaski. http://www.cle. org.pk/Publication/papers/2006/n3117.pdf (accessed 5 December 2014)

#### Baums, Stefan

2006 Towards a computer encoding for Brāhmī. In: Adalbert J. Gail, Gerd J. R. Mevissen, and Richard Salomon (eds.), *Script and image: Papers on art and epigraphy*, 111–143. (Papers of the 12th World Sanskrit Conference 11.1.) Delhi: Motilal Banarsidass.

#### Baums, Stefan

2009 A Gāndhārī commentary on early Buddhist verses: British Library Kharosthī fragments 7, 9, 13 and 18. University of Washington PhD dissertation. Pro-Quest Dissertations 0822059.

#### Baums, Stefan

2011 Indiske skrifter: Seglskrift, aksaras og "arabertal". In: Stig T. Rasmussen, ed., Verdens skrifter, 239–263. København: Forlaget Vandkunsten.

#### Baums, Stefan

2014 Gandhāran scrolls: Rediscovering an ancient manuscript type. In: Jörg Quenzer and Jan U. Sobisch (eds.), *Manuscript cultures: Mapping the field*, 183–225. (Studies in Manuscript Cultures 1.) Berlin: De Gruyter.

#### Baums, Stefan, and Andrew Glass

- Ongoing Catalog of Gāndhārī texts. http://gandhari.org/catalog/ (accessed 12 December 2014)
- Beg, Fazal Amin, John Mock, and Mir Ali Wakhani
  - 2014 Recent developments in Wakhi orthography. Manuscript submitted for publication.
- Bhurgri, Abdul-Majid
  - MS Enabling Pakistani languages through Unicode. http://www.bhurgri.com/ bhurgri/downloads/PakLang.pdf (accessed 5 December 2014)
- Boyer, Auguste M., E. J. Rapson, E. Senart, and P. S. Noble
  - 1920–1929 Kharosthī inscriptions discovered by Sir Aurel Stein in Chinese Turkestan. Oxford: Clarendon Press.

#### Braarvig, Jens (ed.)

2000 *Buddhist manuscripts, volume I.* (Manuscripts in The Schøyen Collection 1.) Oslo: Hermes Publishing.

#### Bright, William

1990 Language variation in South Asia. New York: Oxford University Press.

# Bright, William

1994 Evolution of the Indian writing system. In: Hartmut Günther and Otto Ludwig (eds.), Schrift und Schriftlichkeit: Ein interdisziplinäres Handbuch internationaler Forschung, 1527–1535. (Handbücher zur Sprach- und Kommunikationswissenschaft 10.) Berlin: De Gruyter.

# Bright, William

1999 A matter of typology: Alphasyllabaries and abugidas. *Studies in the Linguistic Sciences* 30(1): 63–71.

#### Bright, William, and Peter T. Daniels

1996 The world's writing systems. New York: Oxford University Press.

#### Brough, John

1977 The arapacana syllabary in the old Lalitāvistara. Bulletin of the School of Oriental and African Studies 40: 85–95.

#### Buddruss, Georg

1982 *Khowar-Texte in arabischer Schrift.* (Akademie der Wissenschaften und der Literatur, Abhandlungen der Geistes- und Sozialwissenschaftlichen Klasse 1.) Wiesbaden: Steiner.

#### Buddruss, Georg

1983 Neue Schriftsprachen im Norden Pakistans: Einige Beobachtungen. In: Aleida Assmann, Jan Assmann, and Christof Hardmeier (eds.), Schrift und Gedächtnis: Beiträge zur Archäologie der literarischen Kommunikation, 231– 244. München: W. Fink.

#### Burnell, Arthur Coke

1874 Elements of South-Indian palæography from the fourth to the seventeenth century A. D.: Being an introduction to the study of South-Indian inscriptions and MSS. 1<sup>st</sup> ed. Mangalore: Stolz & Hirner, Basel Mission Press.

#### Burnell, Arthur Coke

- 1878 Elements of South-Indian palæography from the fourth to the seventeenth century A. D.: Being an introduction to the study of South-Indian inscriptions and MSS. 2<sup>nd</sup> ed. London: Trübner & Co.
- Burushaski Research Academy
  - 2006 Burūšaskī-Urdū luyat, jild-e awwal [Burushaski-Urdu dictionary, vol. 1]. Karachi: Director, Bureau of Composition, Compilation & Translation, University of Karachi.

Burushaski Research Academy

2009 Burūšaskī-Urdū luyat, jild-e doam [Burushaski-Urdu dictionary, vol. 2]. Karachi: Director, Bureau of Composition, Compilation & Translation, University of Karachi.

# Burushaski Research Academy

2011 Burushaski alphabets with examples. http://www.scribd.com/doc/31593787/ Burushaski-Alphabets-With-Examples (accessed 5 Dec. 2014)

#### Burushaski Research Academy

- 2014 Burūšaskī-Urdū luyat, jild-e soam [Burushaski-Urdu dictionary, vol. 3]. Karachi: Director, Bureau of Composition, Compilation & Translation, University of Karachi.
- Cardona, George, and Dhanesh Jain (eds.)
  - 2003 The Indo-Aryan languages. London/New York: Routledge.
- Chengappa, S., S. Bhat, and P. Padakannaya
  - 2004 Reading and writing skills in multilingual/multiliterate aphasics: Two case studies. *Reading and Writing: An Interdisciplinary Journal* 17: 121–135.

#### Chitrali, Rahmat Aziz

2004 Balti primer. Karachi/Chitral: Khowar Academy. http://www.scribd.com/ doc/247229102/Balti-Language-Qaida-by-Rehmat-Aziz-Chitrali-Publishedby-Khowar-Academy-Chitral-Pakistan (accessed 4 December 2014)

# Chrisomalis, Stephen

2010 *Numerical notation: A comparative history*. Cambridge: Cambridge University Press.

Coningham, Robin A. E., F. Raymond Allchin, Catherine M. Batt, and D. Lucy

- 1996 Passage to India? Anuradhapura and the early use of the Brahmi script. *Cambridge Archaeological Journal* 6: 73–97.
- Coulmas, Florian
  - 2003 Writing systems: An introduction to their linguistic analysis. Cambridge: Cambridge University Press.
- Cunningham, Alexander
  - 1875 *Report for the year 1872–73, volume V.* (Archæological Survey of India.) Calcutta: Office of the Superintendent of Government Printing.
- Dani, Ahmad Hasan
  - 1963 Indian palaeography. Oxford: Clarendon Press.
- Dani, Ahmad Hasan

1986 *Indian palaeography.* 2nd ed. New Delhi: Munshiram Manoharlal Publishers. Daniels, Peter T., and William Bright (eds.)

1996 The world's writing systems. New York: Oxford University Press.

#### Das Gupta, Charu Chandra

1958 The development of the Kharosthī script. Calcutta: Firma K. L. Mukhopadhyay. DeSilva, M. W. Sugathapala

1969 The phonological efficiency of the Maldivian writing system. *Anthropological Linguistics* 11: 199–208.

#### Dimitrov, Dragomir

2010 The Bhaikşukā manuscript of the Candrālamkāra: Study, script tables, and facsimile edition. (Harvard Oriental Series 72.) Cambridge, MA: Department of Sanskrit and Indian Studies, Harvard University.

#### Einicke, Katrin

2009 Korrektur, Differenzierung und Abkürzung in indischen Inschriften und Handschriften. (Abhandlungen für die Kunde des Morgenlandes 68.) Wiesbaden: Harrassowitz.

#### Elfenbein, Josef

1983 The Brahui problem again. *Indo-Iranian Journal* 25: 103–132.

#### Falk, Harry

- 1993 Schrift im alten Indien: Ein Forschungsbericht mit Anmerkungen. (ScriptOralia 56.) Tübingen: Gunter Narr.
- Falk, Harry, and Walter Slaje (eds.)
  - 2000–2005 IndoSkript: Eine elektronische indische Paläographie. http://www.indo logie.uni-halle.de/forschung/indoskript/ (accessed 12 December 2014)
- Farmer, Steve, Richard Sproat, and Michael Witzel
  - 2004 The collapse of the Indus-script thesis: The myth of a literate Harappan civilization. *Electronic Journal of Vedic Studies* 11(2): 19–57.

#### Ferguson, Charles A.

- 1978 Patterns of literacy in multilingual situations. In: James E. Alatis (ed.), *International dimensions of bilingual education*, 582–590. Washington: Georgetown University Press.
- Gair, James W., and Bruce D. Cain
  - 1996 Dhivehi writing. In: Daniels & Bright (eds.) 1996: 564–568.

#### Geiger, Wilhelm

1919 *Máldivian linguistic studies.* (Journal of the Ceylon Branch of the Royal Asiatic Society 27.) Colombo: H. C. Cottle, Govt. Printer.

#### Glass, Andrew

2000 *A preliminary study of Kharosthī manuscript paleography.* University of Washington MA thesis.

#### Glück, Helmut

1994 Schriftlichkeit und Diglossie. In: Hartmut Günther and Otto Ludwig (eds.), Schrift und Schriftlichkeit: Ein interdisziplinäres Handbuch internationaler Forschung, 739–766. Berlin: De Gruyter.

#### Gupta, Ashum

2004 Reading difficulties of Hindi-speaking children with developmental dyslexia. *Reading and Writing* 17: 79–99.

#### Hartmann, Jens-Uwe

2000 Zu einer neuen Handschrift des Dīrghāgama. In: Christine Chojnacki, Jens-Uwe Hartmann, and Volker M. Tschannerl (eds.), *Vividharatnakarandaka: Festgabe für Adelheid Mette*, 359–367. Swisttal-Odendorf: Indica et Tibetica Verlag.

#### Heegård, Jan

2000 Linguistic and political aspects of alphabet-making for a threatened language. In: Carl-Erik Lindberg and Steffen Nordahl Lund (eds.), 17th Scandinavian Conference of Linguistics II, 161–176. (Odense Working Papers in Language and Communication 19.) Odense: Institute of Language and Communication, University of Southern Denmark.

#### Hinüber, Oskar von

1979 Die Erforschung der Gilgit-Handschriften: Funde buddhistischer Sanskrit-Handschriften I. (Nachrichten der Akademie der Wissenschaften in Göttingen, 1: Philologisch-historische Klasse, 327–360.) Göttingen: Vandenhoeck & Ruprecht.

#### Hinüber, Oskar von

1990 Der Beginn der Schrift und frühe Schriftlichkeit in Indien. (Abhandlungen der geistes- und sozialwissenschaftlichen Klasse 1989, 11.) Mainz: Akademie der Wissenschaften und der Literatur.

#### Hinüber, Oskar von

2001 Das ältere Mittelindisch im Überblick. (Österreichische Akademie der Wissenschaften, philosophisch-historische Klasse, Sitzungsberichte 467 / Veröffentlichungen der Kommission für Sprachen und Kulturen Südasiens 20.) 2nd ed. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

#### Hladczuk, John, and William Eller

1987 Comparative reading: An international bibliography. (Bibliographies and Indexes in Education 4.) New York: Greenwood Press.

#### Hunzai, Allamah Nasiruddin Nasir

n.d. Burūśo Birkis [Treasures of the Burusho]. Hunza/Gilgit/Karachi: Burushaski Research Academy.

#### Inam Ullah

2004 Lexical database of the Torwali dictionary. Asia Lexicography Conference, Chiangmai, Thailand, 24–26 May, 2004. http://www.fli-online.org/documents/ languages/torwali/Inam\_paper\_2004.pdf (accessed 5 December 2014)

#### Inam Ullah

2010 *Torwālī-Urdū luyat* [Torwali-Urdu Dictionary]. Lahore: Center for Research in Urdu Language Processing, National University for Computer and Emerging Sciences. http://www.cle.net.pk/otd/ (accessed 5 December 2014)

#### Inam Ullah

n.d. Scripts and characters used in the languages of Northern Pakistan. http://www. cle.org.pk/IDN/download/NPLang-Charachters.pdf (accessed 5 December 2014)

#### Jahani, Carina

- 1989 Standardization and orthography in the Balochi language. (Acta Universitatis Upsaliensis, Studia Iranica Upsaliensia 1). Uppsala: Uppsala University.
- Jahani, Carina, and Agnes Korn
  - 2009 Balochi. In: Gernot Windfuhr (ed.), *The Iranian languages*, 634–692. London/ New York: Routledge.

#### Jain, Dhanesh

2003 Sociolinguistics of the Indo-Aryan languages. In: Cardona & Jain (eds.) 2003: 391–443.

#### Janert, Klaus Ludwig

1972 Abstände und Schlußvokalverzeichnungen in Aśoka-Inschriften: mit Editionen und Faksimiles in 107 Lichtdrucktafeln. (Verzeichnis der orientalischen Handschriften in Deutschland, Supplementband 10.) Wiesbaden: Steiner.

#### Jensen, Hans

- 1969 Die Schrift in Vergangenheit und Gegenwart. Berlin: Deutscher Verlag der Wissenschaften.
- Jha, N., and Navaratna Srinivasa Rajaram
  - 2000 The deciphered Indus script: Methodology, readings, interpretations. New Delhi: Aditya Prakashan.
- Joshi, Jagat Pati, and Asko Parpola (eds.)
  - 1987 Corpus of Indus seals and inscriptions 1: Collections in India. (Annales Academiae Scientiarum Fennicae B 239 / Memoirs of the Archaeological Survey of India 86.) Helsinki: Suomalainen Tiedeakatemia.
- Karanth, Pratibha, and M. G. Suchitra
  - 1993 Literacy acquisition and grammaticality judgments in children. In: Robert J. Scholes (ed.), *Literacy and language analysis*,143–156. Hillsdale: Lawrence Erlbaum Associates.
- Karanth, Pratibha, Anu Mathew, and Priya Kurien
  - 2004 Orthography and reading speed: Data from native readers of Kannada. *Reading and Writing* 17: 101–120.
- Kazmi, Syed Muhammad Abbas
  - 1996 The Balti language. In: P. N. Pushp and K. Warikoo (eds.), *Jammu, Kashmir & Ladakh: Linguistic predicament*. New Delhi: Himalayan Research and Cultural Foundation/Har Anand Publications. http://koshur.org/Linguistic/7. html (accessed 4 December 2014)

#### Kenoyer, Jonathan Mark

1998 Ancient cities of the Indus civilization. Karachi: Oxford University Press.

# Khan, Tarik Ali

2002 Little Tibet: Renaissance and resistance in Baltistan. http://www.phayul.com/ news/article.aspx?id=325&t=0 (accessed 4 December 2014)

#### Khubchandani, Lachman M.

2003 Sindhi. In: Cardona & Jain (eds.) 2003: 622–658.

#### Knozorov, Yurii (ed.)

1965 Predvaritel'naja soobščenie ob issledovanii protoindijskix tekstov [Preliminary communication on the study of the proto-Indian texts]. Moskva: Akademia Nauk SSSR.

#### Koul, Omkar N.

1995 Standardization of Kashmiri script. In: Imtiaz Hasnain (ed.), Standardization and modernization: Dynamics of language planning, 269–278. New Delhi: Bahri.

#### Lamuwal, Abd-El-Malek, and Adam Baker

2013 Southeastern Pashayi. Journal of the International Phonetic Association 43(2): 243–246. http://journals.cambridge.org/action/displayJournal?jid=IPA (accessed 3 December 2014)

#### Losey, Wayne E.

2002 Writing Gojri: Linguistic and sociolinguistic constraints on a standardized orthography for the Gujars of South Asia. University of North Dakota MA thesis. http://www.fli-online.org/documents/languages/gojri/losey\_thesis.pdf (accessed 3 December 2014)

#### MacKenzie, D. N.

1959 A standard Pashto. Bulletin of the School of Oriental and African Studies 22(1/3): 231–235.

# MacKenzie, D. N.

1997 The development of the Pashto script. In: Shirin Akiner and Nicholas Sims-Williams (eds.), *Languages and scripts of Central Asia*, 137–143. London: School of Oriental and African Studies.

#### Mahadevan, Iravatham

1977 *The Indus script: Texts, concordance and tables.* (Memoirs of the Archaeological Survey of India 77.) New Delhi: Archaeological Survey of India.

#### Mahadevan, Iravatham

2003 Early Tamil epigraphy: From the earliest times to the sixth century A. D. (Harvard Oriental Series 62.) Chennai: Cre-A.

# Malik, M. G. Abbas

2005 *Towards a Unicode compatible Punjabi character set*. 27th Internationalization and Unicode Conference, April 2005, Berlin. https://hal.inria.fr/file/index/ docid/1002347/filename/mgam05–1.pdf (accessed 24 December 2014)

#### Malmquist, Eve (ed.)

1982 Handbook on comparative reading: An annotated bibliography and some viewpoints on university courses in comparative reading. Newark: International Reading Association.

#### Maurer, Walter H.

1976 On the name Devanāgarī. *Journal of the American Oriental Society* 96: 101–104. Mirdehghan, Mahinnaz

2010 Persian, Urdu, and Pashto: A comparative orthographic analysis. Writing Systems Research 2(1): 9–23.

#### Mock, John Howard

1998 The discursive construction of reality in the Wakhi community of Northern Pakistan. University of California PhD dissertation. ProQuest Dissertations 9922976. Namus, M. Shuja

1961 *Gilgit aur šinā zabān* [Gilgit and the Shina language]. Bahawalpur, Pakistan: Urdu Academy.

National Language Authority, Pakistan

2002 Balti qāeda [Balti primer]. Islamabad: National Language Authority.

Norris, Edwin

1846 On the Kapur-di-Giri rock inscription. *The Journal of the Royal Asiatic Society* of Great Britain and Ireland 8: 303–307.

Oommen, Chinna

1973 India. In: John A. Downing (ed.), Comparative reading: Cross-national studies of behavior and processes in reading and writing, 403–425. New York: Macmillan.

Pandey, Anshuman

2010 Introducing another script for writing Balti. http://www-personal.umich. edu/~pandey/ (accessed 4 December 2014)

#### Parpola, Asko

1994 Deciphering the Indus script. New York: Cambridge University Press.

Parpola, Asko

1996 The Indus script. In: Daniels & Bright (eds.) 1996: 165–171.

Parpola, Asko

2008 Is the Indus script indeed not a writing system? In: Airāvati: Felicitation volume in honour of Iravatham Mahadevan, 111–131. Chennai: Varalaaru. com.

Parpola, Asko, B. M. Pande, and Petteri Koskikallio (eds.)

2010 Corpus of Indus seals and inscriptions 3: New material, untraced objects and collections outside India and Pakistan 1: Mohenjo-Daro and Harappa. (Annales Academiae Scientiarum Fennicae B 359 / Memoirs of the Archaeological Survey of India 96.) Helsinki: Suomalainen Tiedeakatemia.

#### Parvez, Aslam

1996 *The adaptation of Perso-Arabic script for Urdu, Panjabi, and Sindhi.* New Delhi: Monumental Publishers.

Patel, Purushottam G.

1993 Ancient India and the orality-literacy divide theory. In: Robert J. Scholes (ed.), *Literacy and language analysis*, 199–208. Hillsdale: Lawrence Erlbaum Associates.

Patel, Purushottam G.

1995 Brahmi scripts, orthographic units and reading acquisition. In: Taylor & Olson (eds.): 265–275.

Pederson, Eric

2003 Mirror-image discrimination among nonliterate, monoliterate, and biliterate Tamil subjects. *Written Language and Literacy* 6(1): 71–91.

# Penzl, Herbert

1954 Orthography and phonemes in Pashto (Afghan). Journal of the American Oriental Society 74(2): 74–81.

# Possehl, Gregory L.

1996 Indus age: The writing system. Philadelphia: University of Pennsylvania Press.

Prakash, P., and R. Malatesha Joshi

- 1995 Orthography and reading in Kannada: A Dravidian language. In: Taylor & Olson (eds.) 1995: 95–108.
- Prakash, P., D. Rekha, R. Nigam, and P. Karanth
  - 1993 Phonological awareness, orthography, and literacy. In: Robert J. Scholes (ed.), *Literacy and language analysis*, 55–70. Hillsdale: Lawrence Erlbaum Associates.

#### Prinsep, James

1837 Note on the facsimiles of inscriptions from Sanchí near Bhilsa, taken for the Society by Captain Ed. Smith, Engineers. *The Journal of the Asiatic Society of Bengal* 6: 451–477.

#### Prinsep, James

1838 Additions to Bactrian numismatics, and discovery of the Bactrian alphabet. The Journal of the Asiatic Society of Bengal 7: 636–658.

#### Rao, S. R.

1982 The decipherment of the Indus script. Bombay: Asia Publishing.

#### Rasoolpuri, Aslam

1976 Sirāikī rasmulxat kī muxtasar tārīx [A short history of Siraiki orthography]. Multan: Bazme Saqafat.

#### Sagar, Muhammd Zaman

2008 A multilingual education project for Gawri-speaking children in northern Pakistan. 2<sup>nd</sup> International Conference on Language Development, Language Revitalization, and Multilingual Education in Ethnolinguistic Communities, Bangkok 1–3 July, 2008. http://www.seameo.org/\_ld2008/doucments/ Presentation\_document/Gawri\_presentation\_Bangkok2008.pdf (accessed 24 December 2014)

#### Sakhi, Ahmad Jami

2000 waxī zabān – tārīx ke āene mē – mah qāeda [Wakhi language in the mirror of history, including a primer]. Gilgit: Ahmad Jami Sakhi, printed at Farman Printing Press.

#### Salomon, Richard

1987 A recent claim to decipherment of the "shell script". Journal of the American Oriental Society 107: 313–315.

#### Salomon, Richard

1995 On the origin of the early Indian scripts. *Journal of the American Oriental* Society 115: 271–279.

#### 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.

#### Salomon, Richard

1999 Ancient Buddhist scrolls from Gandhāra: The British Library Kharosthī fragments. Seattle: University of Washington Press.

#### Salomon, Richard

2000 Typological observations on the Indic script group and its relationship to other alphasyllabaries. *Studies in the Linguistic Sciences* 30: 87–103.

Salomon, Richard

2008 Whatever happened to Kharosthī? The fate of a forgotten Indic script. In: John Baines, John Bennet, and Stephen Houston (eds.), *The disappearance* of writing systems: Perspectives on literacy and communication, 139–155. London: Equinox Publishing.

#### Sander, Lore

- 1968 Paläographisches zu den Sanskrithandschriften der Berliner Turfansammlung. (Verzeichnis der orientalischen Handschriften in Deutschland, Supplementband 8.) Wiesbaden: Steiner.
- Sander, Lore
  - 2000 A brief paleographical analysis of the Brāhmī manuscripts in volume I. In: Jens Braarvig (ed.), *Buddhist manuscripts*, volume I, 285–300. Oslo: Hermes Publishing.
- Schmidt, Ruth Laila, and Razwal Kohistani
  - 1995 The mirror of writing. Third International Hindu Kush Cultural Conference, 26–30 August, 1995 in Chitral, Pakistan. Original and more accurate version at http://www.hf.uio.no/ikos/english/research/projects/shina/publications/ Schmidt%20and%20Kohistani%202008-original-1.pdf (accessed 3 December 2014)

#### Sebba, Mark

- 2007 Spelling and society: The culture and politics of orthography around the world. Cambridge: Cambridge University Press.
- Sebba, Mark, Shahrzad Manootian, and Carla Jonsson (eds.)
  - 2012 Language mixing and code-switching in writing: Approaches to mixedlanguage written discourse. New York/Abingdon, Oxon: Routledge.

#### Shackle, Christopher

- 2003 Panjabi. In: Cardona & Jain (eds.) 2003: 581–621.
- Shah, Sayid Ghulam Mustafa, and Asko Parpola (eds.)
  - 1991 Corpus of Indus seals and inscriptions 2: Collections in Pakistan. (Annales Academiae Scientiarum Fennicae B 240 / Memoirs of the Department of Archaeology and Museums, Government of Pakistan 5.) Helsinki: Suomalainen Tiedeakatemia.

# Sircar, Dines Chandra

1942 Select inscriptions bearing on Indian history and civilization. Calcutta: University of Calcutta.

#### Sircar, Dines Chandra

1965a Indian epigraphy. Delhi: Motilal Banarsidass.

#### Sircar, Dines Chandra

1965b Select inscriptions bearing on Indian history and civilization, Volume I: From the sixth century B. C. to the sixth century A. D. 2<sup>nd</sup> ed. Calcutta: University of Calcutta.

# Sircar, Dines Chandra

1966 Indian epigraphical glossary. Delhi: Motilal Banarsidass.

# Sircar, Dines Chandra

1983 Select inscriptions bearing on Indian history and civilization, Volume II: From the sixth to the eighteenth century A. D. Delhi: Motilal Banarsidass.

Sprigg, Richard K.

1996 My Balti-Tibetan and English dictionary. *The Tibet Journal* 21(4): 3–22. Sprigg, Richard K.

2002 Balti-English, English-Balti dictionary. London: Routledge Curzon.

#### Sproat, Richard

2000 A computational theory of writing systems. (AT&T Labs research report.) http:// www.cslu.ogi.edu/~sproatr/newindex/wsbook.pdf (accessed 6 December 2014)

#### Sproat, Richard

2006a Brahmi-derived scripts, script layout, and segmental awareness. Written Language and Literacy 9(1): 45-65.

#### Sproat, Richard

2006b *A computational theory of writing systems*. Cambridge: Cambridge University Press.

#### Steinkellner, Ernst

2004 *A tale of leaves: On Sanskrit manuscripts in Tibet, their past and their future.* Amsterdam: Royal Netherlands Academy of Arts and Sciences.

#### Strauch, Ingo

2008 The Bajaur collection of Kharosthī manuscripts: A preliminary survey. *Studien zur Indologie und Iranistik* 25: 103–36.

#### Strauch, Ingo

2012 The character of the Indian Kharoṣṭhī script and the "Sanskrit revolution": A writing system between identity and assimilation. In: Alex de Voogt and Joachim Friedrich Quack (eds.), *The idea of writing: Writing across borders*, 131–168. Leiden: Brill.

#### Swank, Heidi

2008 It all hinges on the vowels: Reconsidering the alphasyllabary classification. *Written Language and Literacy* 11: 73–89.

#### Taj, Abdul Khaliq

1989 *Şiņā qā 'eda* [Shina primer]. Gilgit: Usmani Kitabkhana.

#### Taylor, Insup, and David R. Olson

1995 Scripts and literacy: Reading and learning to read alphabets, syllabaries and characters. (Neuropsychology and Cognition 7.) Dordrecht: Kluwer.

### Trail, Ron, and Gregory Cooper

1999 Kalasha dictionary – with English and Urdu. (Studies in Languages of Northern Pakistan 7.) Islamabad: National Institute of Pakistan Studies, Quaidi-Azam University/Summer Institute of Linguistics. http://www.fli-online.org/ (accessed 6 December 2014)

#### Ul-Mulk, Samsam

n.d. *Khovār qā'eda* [Khowar primer]. Peshawar: Department of Publications, NWFP.

#### Unicode Consortium

1991– http://unicode.org/ (accessed 12 December 2014)

#### Unicode Consortium

2011 Arabic Supplement code page. http://www.unicode.org/charts/ (accessed 8 January 2015)

#### Unseth, Peter

- 2005 Sociolinguistic parallels between choosing scripts and languages. Written Language and Literacy 8: 19–42.
- Vaid, Jyotsna
  - 1995 Script directionality affects nonlinguistic performance: Evidence from Hindi and Urdu. In: Taylor & Olson (eds.) 1995: 295–310.

van Schaik, Sam

 A new look at the Tibetan invention of writing. In: Yoshiro Imaeda, Matthew T. Kapstein, and Tsuguhito Takeuchi (eds.), New studies of the Old Tibetan documents: Philology, history and religion, 45–96. Tokyo: Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.

#### Vasanta, Duggirala

- 2004 Processing phonological information in a semi-syllabic script: Developmental data from Telugu. *Reading and Writing* 17: 59–78.
- Veldhuis, Dorina, and Jeanne Kurvers
  - 2012 Offline segmentation and online language processing units. Written Language & Literacy 15(2): 165–184.
- Wali, Aamir, Richard Sproat, Prakash Padakannaya, and Bhuvaneshwari B.
  - 2009 Model for phonemic awareness in readers of Indian script. *Written Language* and Literacy 12: 161–169.

#### Weingarten, Rüdiger

2011 Comparative graphematics. *Written Language and Literacy* 14: 12–38.

# Wheeler, Mortimer

1968 The Indus civilization: Supplementary volume to the Cambridge history of India. 3<sup>rd</sup> ed. Cambridge: Cambridge University Press.

#### Yadav, Nisha, and M. N. Vahia

2011 Indus script: A study of its sign design. *Scripta* 3: 1–36.

Yun, Ju-Hong

2003 Pashai language development project: Promoting Pashai language, literacy and community development. Conference on Language Development, Language Revitalization and Multilingual Education Activities, Bangkok, Thailand, 6–8 November 2003. http://www.sil.org/asia/ldc/parallel\_presentations.html (accessed 6 December 2014)

# Zia, Mohammad Amin

1986 *Şiņā qāeida aur græmar* [Shina primer and grammar]. Gilgit (Pakistan): Zia Publications.

#### Zia, Mohammad Amin

2010 *Şiņā-Urdū luyat* [Shina-Urdu dictionary]. Gilgit (Pakistan): Zia Publications. Zide, Norman

1996 Scripts for Munda languages. In: Daniels & Bright (eds.) 1996: 612–618.

# The World of Linguistics

# *Editor* Hans Henrich Hock

Volume 7

De Gruyter Mouton

# The Languages and Linguistics of South Asia

A Comprehensive Guide

*Edited by* Hans Henrich Hock Elena Bashir

De Gruyter Mouton

ISBN 978-3-11-042715-8 e-ISBN (PDF) 978-3-11-042330-3 e-ISBN (EPUB) 978-3-11-042338-9

*Library of Congress Cataloging-in-Publication Data* A CIP catalog record for this book has been applied for at the Library of Congress.

Bibliographic information published by the Deutsche Nationalbibliothek The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.d-nb.de.

© 2016 Walter de Gruyter GmbH, Berlin/Boston Cover image: Chogolisa, Karakorum/Uwe Steffens/ullstein bild Typesetting: Dörlemann Satz GmbH & Co. KG, Lemförde Printing and binding: CPI books GmbH, Leck © Printed on acid-free paper

Printed in Germany

www.degruyter.com

