

THE TRANSLITERATION
OF
ORIENTAL ALPHABETS.
BY
JAMES BURGESS.

THE TRANSLITERATION
OF
ORIENTAL ALPHABETS.
BY
JAMES BURGESS.

The transliteration of Oriental alphabets by means of the Roman letters has frequently occupied the attention of scholars since the time of Sir William Jones, whose paper on the subject stands first in the first volume of the *Asiatic Researches*. But he was not the first to propose a systematic method of spelling Eastern words; for, as he mentions, Mr. H. B. Halhead had published, in 1775, a scheme, on which Charles Wilkins had improved, and to which he adhered in his papers subsequently.

The system of Jones as applied to the Sanskritic alphabets, has since been modified by Wilson and others, but continues substantially to be the system still in use. His proposals for the Arabic alphabet did not receive the same attention, and different scholars having employed systems of their own, there is now more diversity in the transliteration of Arabic, Turkish, Persian, and Hindustani than in that of the Sanskritic alphabets.

If any one system of transliteration of Oriental alphabets is now to have a chance of general acceptance, it is manifest that it must take into account something more than the consensus of specialists, and must be adapted to the wants of a far larger class than that of philologists: that is, to geogra-

phers, cartographers, travellers, writers on all sorts of Oriental subjects, teachers, and the like, as well as to Indian officers and the vast and increasing number of Hindus using the English language and alphabet. The great Geographical Societies and the Hydrographic offices of commercial nations are also seeking after a uniform and rational system for the spelling of place-names in the East¹). What an advantage it would be at the present time to have all these enlisted on the side of one practical and fairly consistent method! The battle would be almost won if scholars could carry with them the classes alluded to; and to do this they must yield on some small points and agree to accept one system among themselves.

The general considerations for guidance, to which all must agree, as conditions of the whole problem, are as follow: —

1. As we mean transliteration into the Roman alphabet only, we are not at liberty to add fresh signs, such as Greek letters or modified Roman type²) other than by the addition of simple diacritical marks.

2. While it is desirable to represent each Oriental character by a single Roman letter, the paucity of letters in the latter alphabet renders it necessary to use combinations of two letters, or digraphs, for aspirated sounds and in cases for which the Roman alphabet has no single letter of similar sound.

3. Such letters only should be used with diacritical marks as, in case of the latter being omitted (as in maps and general literature), the pronunciation of the words will not, to European ears, be materially affected. Jones's *cā* for the first Sanskrit palatal, and *c'h* for the second guttural, made words like *chaṇḍa* and *c'haṇḍa* (*khaṇḍa*), or *suchā* and *suc'hā* (*sukhā*), when printed without the diacritical marks, absolutely alike.

4. Transliteration should be based on the classification of the sounds of the language; and whatever system is arranged for one family of languages should apply to any other with

1) See *Report of the United States Hydrographic Office Board on the Orthography of Geographic-Names* (Washington, 1891); *Bulletin Soc. Géogr. Paris*, 1886, pp. 198—202; *Annalen der Hydrographie*, 1888, p. 548 f.; *Nachrichten für Seefahrer*, 1888, p. 624 f.; &c.

2) Conf. Whitney in *Journ. Am. Orient. Soc.*, vol. VII, p. 339.

only necessary modifications. That is, whatever diacritical marks are used to distinguish one class of sounds in one language, the same marks should be applied as far as necessary to the same class in any other¹⁾. Also, the same letter should represent the sound in a second language, nearest to that for which it stands in the language for which the system has first been arranged.

5. We should utilize as far as practicable, the diacritical marks already in use, rather than introduce new sets of signs.

6. In India, that is over the whole Sanskritic area, English spelling must necessarily exercise a predominating influence. The civil and military officials, the educated Hindus, everyone in fact in the East, will adopt English, rather than Continental forms of spelling. They cannot be expected to accept, for example, the soft *g* as the transcript of the medial palatal ङ, or *k* for the tenuis; such a proposal, under existing circumstances, would only court failure. In any such case „it is wiser for Muhammad to go to the hill“. The success or failure of any scheme must depend very largely on its taking into consideration this controlling fact, in all its bearings.

7. Whilst the common pronunciation is necessarily taken into account, in selecting the letters of one alphabet to represent those of another, it is to be remembered that *transcription* does not attempt to indicate differences of pronunciation. A letter may be hard in one position and soft in another, and the sound of its transcript must be subject to the same rule as in its proper character. All variations of pronunciation must be learnt from the grammar or from natives.

THE INDIAN ALPHABETS.

Most of the alphabets of India correspond with the *devanāgarī*, the transcription of which, except for a few letters, may be said to be now fixed.

To mark the long vowels (*ā*, *ī*, *ū*), — especially in Sanskrit texts, where accents have to be attended to, — the *macron* is the natural mark and therefore to be preferred; but in maps and

1) Conf. Lepsius's *Standard Alphabet*, 2nd. ed., pp. 59 ff.

general literature, the circumflex (more frequent in English founts) leads to no inconvenience. The *guṇa* diphthongs *e* and *o* require no *macron*, as there is only one long sound for each; and the mark should be reserved for the longer forms in the Dravidian alphabets.

For the *anusvāra*, *m* with a dot above or below is the alternative. The dot below is distinctive of the linguals, but if placed above, it distinguishes this sound from them, while it corresponds in position to the *devanāgarī* sign, and being already largely in use, *m̄* is distinctly preferable ¹⁾.

Among the nasals, the lingual by analogy takes the dot below; naturally *ṅ* (provided in most founts) answers for the palatal ²⁾; and *ṅ̄*, already so much used for it, is most suitable for the guttural, and, like the *anusvāra*, it bears some analogy to the *Devanāgarī* letter with its dot. To place the dot after the letter (*ṅ̄*) is awkward in printing; and to use a new sign (*ṅ̄*), not in the Roman alphabet, is quite against the conditions of the problem.

For the surd palatals, Jones, Wilson, and others have employed *cā* and *cā̄*, — the latter on analogy, as the aspirated form of the first. Objection is made to the use of *cā* that it is a compound sign for a simple sound. But this is frequent in all languages using the Roman letters and misleads nobody, — least of all scholars. The sound, in European languages, is almost peculiar to the English, Russian and Italian; and *cā*, though written in two letters, is as much a simple sound in English as it is in Sanskrit ³⁾. We also use digraphs for at least ten other single Sanskrit letters. Mr. Crow, a missionary in Bengal, proposed the use of *c* alone for this sound, which is purely a conventional use of the letter; and though it may be easy for specialists to attach an arbitrary sound to a letter, the public can hardly be expected to recognize the sounds intended when we write *cakra*, *Cicoli*, *candra*, *cacandra*, *Cenclū*, *Cacara*, &c. As usual in all European languages, the *c* will be sounded

1) So Whitney, *Proc. Am. Orient. Soc.*, vol. XI, p. LXX.

2) The palatal and palato-dental letters are marked by scientific phonologists, following Bopp, *above* the characters, when any mark is required.

3) See Lepsius, *Standard Alphabet*, 2nd. ed., p. 8.

hard before broad vowels and consonants ¹⁾. Thus to upset use and wont everywhere outside the small circle of Sanskritists, is impossible: the latter know quite as well as the public what *cā* is meant to represent, and may as well yield so small a point. Every map and geographical text-book ²⁾ has adopted *cā*, as does the roll of every Indian regiment ³⁾ and every revenue record, and we may as well follow. *Cāā* is the analogous aspirate of this, unless we adopt for it *čā* (with a diacritical mark); the first is no more clumsy than many combinations in German, English, and French, — as *tsch*, &c.

It is difficult to represent differently the three Indian sibilants. Jones proposed *ś* (with an accent) for the palatal, and *ṣ* for the lingual sibilants. Wilson and Thomas substituted *ṣ* (with a dot below) for the palatal, assimilating it to the linguals, to which Whitney very justly objected „as against every analogy and altogether to be condemned“. Continental scholars have introduced *ç* (with the *cedilla*) for it. Not being a letter of the Roman alphabet, nothing can well be said in its favour, except that it is a type found in French founts, introduced as a makeshift to represent a soft *c* before a broad vowel, where the Italians would insert an *i*. In India it has little chance of general use, nor by cartographers and in general literature. To affix a *cedilla* to *ç* (*ç*), as has been suggested, is introducing a type in no alphabet and altogether objectionable as contravening the third guiding consideration. We distinguish the palatal nasal *ñ* by a *tilde* over the letter: might we not distinguish this palatal also by the same mark as *ñ*? Or, can we not go back on the popular and analogous forms for the two palatals, viz.: *śā* ⁴⁾ and *cā*?

1) Even in Italian an *i* is inserted before a broad vowel to give the soft sound to *c*, as in *cio*.

2) The Government of India, in ordering that the names in the official *Gazetteers* should be spelt in accordance with Wilson's revised form of Sir. W. Jones's system of transliteration, gave an impulse to Indian orthography more influential in an indirect way than is perhaps quite realized. It has found its way into geographical text-books such as Constable's admirable *Hand-Atlas of India*, where it is presented to both student and tourist. See *Scot. Geog. Mag.*, vol. VII (1891), pp. 857 f.; vol. VIII, pp. 23 f.

3) Conf. C. J. Lyall's *Guide to the Transliteration of Hindu and Muhammadan Names in the Bengal Army*, 3rd. ed., 1892.

4) The French Geographical Society has adopted *śā* to represent the French sound of *cā* in foreign place-names; and this has been authorized in the French Hydrographic service.

For the lingual, *ṣ* has been most largely used hitherto; *ṣ̣* (with a dot below) was advanced by Grassmann and others; but, — while analogy requires the diacritical mark, — when printed without it, the indication of the proper sound will be entirely lost. The objection to a digraph disappears in presence of the representation of the aspirates, and *ṣ̣* (with the diacritical mark ') of the linguals) best represents this sound.

Lastly, the lingual vowels *r̄* and *ḷ* are properly marked as the lingual consonants, and while it is scientifically true that, for the scientist, they do not require a vowel sign, yet — for the public — forms like *r̄ṣi*, *r̄kṭha*, *Kṛṣṇa*, &c. appear unpronounceable, and custom has long familiarized us with *r̄i* or *r̄ī*, *ḷī*, &c. And with the *ī* marked *short*, the merest tyro could not be misled in transliterating back into *devanāgarī*; and *r̄īṣhi*, *r̄īkṭha*, *Kṛṣṇa*, would be pronounced with some approach to their proper sounds. To mark the *r̄* with a *macron*, when we cannot conveniently treat the *ḷ* similarly, is not satisfactory. For the rarer long forms, we may write *r̄ī* and *ḷī*. In Telugu, Kanarese, and Malayalam, we have also the lingual consonants *ḷ* and *r̄*, for which we should reserve the single letters.

The accompanying table will shew at a glance the principal Indian alphabets, with the transliterations suggested. In the column headed „Hindustānī“, however, the whole alphabet is not given in the first table, but only in the second.

1) The dot under the letter to mark a cerebral lingual was introduced by Bopp and his school.

FIRST TABLE.

Xth Congrès international des Orientalistes. — Section I. (Pages 33—36.)

THE PRINCIPAL INDIAN ALPHABETS.

Roman	Devanāgarī	Bengali	Uriya	Telugu	Kanarese	Tamil	Grantha	Malayālam	Singha- lese	Hindu- stānī
a	अ	অ	ଅ	అ	ಅ	அ	𑀅	അ	𑀅	ا
ā	आ	আ	ଆ	ఆ	ಆ	ஆ	𑀆	ആ	𑀆	آ
i	इ	ই	ଇ	ఇ	ಇ	ஐ	𑀇	ഈ	𑀇	ی
ī	ई	ঐ	ଐ	ఱ	ಱ	ற	𑀈	ഐ	𑀈	یٰ
u	उ	উ	ଉ	ఉ	ಁ	உ	𑀉	ഉ	𑀉	و
ū	ऊ	ঊ	ଊ	ఱ	ಱ	ఱ	𑀊	ഊ	𑀊	وٰ
rī	ऋ	ঋ	ଠ	ఱ	ఱ	—	—	ഠ	𑀋	—
ṛī	ॠ	ॠ	ॠ	ఱ	ఱ	—	—	ॠ	𑀌	—
lī	ऌ	ॡ	—	ॡ	ॡ	—	—	ॡ	𑀍	—
ḷī	ॢ	ॢ	—	ॢ	ॢ	—	—	ॢ	𑀎	—
e	ए	এ	ଏ	ఎ	ఎ	எ	𑀏	എ	𑀏	ه
ē	—	—	—	ఎ	ఎ	ஏ	—	—	—	هٰ
ai	ऐ	ঐ	ঐ	ఱ	ఱ	ஐ	𑀐	ഐ	𑀐	هٰ
o	ओ	ঔ	ঔ	ఱ	ఱ	ஔ	𑀑	ഔ	𑀑	و
ō	—	—	—	ఱ	ఱ	—	—	—	—	—
au	औ	ঔ	ঔ	ఱ	ఱ	ఱ	𑀒	ഔ	𑀒	وٰ
m	—	—	—	—	—	—	—	—	—	—
h	ः	—	—	—	—	—	—	—	—	—
k	क	ক	କ	క	ಕ	க	𑀓	ക	𑀓	ك
kh	ख	খ	ଖ	ఖ	ఖ	க	𑀔	ഖ	𑀔	خ
g	ग	গ	ଗ	గ	ಗ	க	𑀕	ഗ	𑀕	گ
gh	घ	ঘ	ଘ	ఘ	ఘ	க	𑀖	ഘ	𑀖	غ
ñ	ङ	ঙ	ଙ	ఙ	ఙ	ங	𑀗	ങ	𑀗	—
ch	च	চ	ଚ	చ	చ	ச	𑀘	ച	𑀘	—
chh	छ	ছ	ଛ	ఱ	ఱ	ச	𑀙	ച	𑀙	—
j	ज	জ	ଜ	జ	ಜ	ச	𑀚	ജ	𑀚	—
jh	झ	ঝ	ଝ	ఱ	ఱ	ச	𑀛	ജ	𑀛	—
ñ	ञ	ঞ	ଞ	ఱ	ఱ	—	—	—	—	—

THE PRINCIPAL INDIAN ALPHABETS.

Roman	Devan-gari	Bengali	Uriya	Telugu	Kanarese	Tamil	Grantha	Malay-alam	Singha-lese	Hindu-stani
t	ट	ট	ठ	ఠ	ಠ	ட	ṭ	ṭ	ṭ	ṭ
th	ठ	ঠ	ॠ	ఠ	ఠ	ఠ	ṭ	ṭ	ṭ	ṭ
d	ड	ড	ॡ	ఢ	ఢ	ఢ	ḍ	ḍ	ḍ	ḍ
dh	ढ	ঢ	ॢ	ఢ	ఢ	ఢ	ḍ	ḍ	ḍ	ḍ
n	ण	ণ	ॣ	ॣ	ॣ	ణ	ṇ	ṇ	ṇ	ṇ
t	न	ন	।	।	।	త	ṅ	ṅ	ṅ	ṅ
th	थ	থ	॥	॥	॥	త	ṅ	ṅ	ṅ	ṅ
d	द	দ	॥	॥	॥	త	ṅ	ṅ	ṅ	ṅ
dh	ध	ধ	॥	॥	॥	త	ṅ	ṅ	ṅ	ṅ
n	म	ম	॥	॥	॥	త	ṅ	ṅ	ṅ	ṅ
p	प	প	॥	॥	॥	త	ṅ	ṅ	ṅ	ṅ
ph	फ	ফ	॥	॥	॥	త	ṅ	ṅ	ṅ	ṅ
b	ब	ব	॥	॥	॥	త	ṅ	ṅ	ṅ	ṅ
bh	भ	ভ	॥	॥	॥	త	ṅ	ṅ	ṅ	ṅ
m	म	ম	॥	॥	॥	త	ṅ	ṅ	ṅ	ṅ
y	य	য	॥	॥	॥	త	ṅ	ṅ	ṅ	ṅ
r	र	র	॥	॥	॥	త	ṅ	ṅ	ṅ	ṅ
l	ल	ল	॥	॥	॥	త	ṅ	ṅ	ṅ	ṅ
ll	ळ	ळ	॥	॥	॥	త	ṅ	ṅ	ṅ	ṅ
ll	ॡ	ॡ	॥	॥	॥	త	ṅ	ṅ	ṅ	ṅ
v	व	ব	॥	॥	॥	త	ṅ	ṅ	ṅ	ṅ
sh, sh	श	শ	॥	॥	॥	త	ṅ	ṅ	ṅ	ṅ
sh	ष	ষ	॥	॥	॥	త	ṅ	ṅ	ṅ	ṅ
s	स	স	॥	॥	॥	త	ṅ	ṅ	ṅ	ṅ
h	ह	হ	॥	॥	॥	త	ṅ	ṅ	ṅ	ṅ
ksh	क्ष	ক্ষ	॥	॥	॥	త	ṅ	ṅ	ṅ	ṅ
n	॥	॥	॥	॥	॥	త	ṅ	ṅ	ṅ	ṅ

THE ARABIC, TURKISH, PERSIAN, & HINDUSTANI
ALPHABETS.

Passing from the Sanskritic to the Arabic family of alphabets, the letters adopted for the one must also stand for the sounds most nearly allied to them in the other. Transcription, as already remarked, does not always carry with it pronunciation. In Turkish, for example, *kāf*, whether sounded as *k* or *y*, is written with the same Arabic letter, and so it should be transliterated; its sound has to be learnt from its position. So with the Arabic *jīm* (ج): it is hard or soft dialectically, but should be always transliterated by the same Roman type, and the sound applied as in reading the native character.

In these alphabets about seventeen letters may be regarded as accepted by all for simple transcription into Roman letters. In Hindustani there are three cerebrals, ت, د, and ر, which, by analogy with the Sanskrit, are represented as *t*, *d*, and *r*. Allied to these are the four gutturo-dentals *sād*, *zād*, *tā*, and *zā*; these might have been quite suitably represented¹⁾, in Lepsius's notation, as *s*, *d*, *t*, *z*, but a majority of scholars have already adopted *s* and *d* (with *z* in Persian and Hindustani) for the first two, and *t*, *z* have also been used for the two latter rather extensively. For Arabic these may be accepted; in the other tongues we have both *sād* and *zā* to be represented by *s*, and we must distinguish between them by a double dot under one of them. If we so mark the *zā*, we may also similarly distinguish the *tā* in Hindustani from its special cerebral, as has been done by Jones and others in India.

Among the dentals, we have to provide for the two lipped letters ث (*sē*) and ذ (*zāl*). The diacritical marks should differ from those of the linguals, and both should be marked alike. The *zāl* has most frequently been marked by a line below the letter, and *sē* occasionally so; hence this may most conveniently indicate these two dentals.

For the palatals — چ, ش, and ژ, — as in Sanskrit, practical considerations must justify the employment of the digraphs,

1) *Journ. Amer. Or. Soc.*, vol. VII, p. 325.

ch, *sh*, and *zh*. To distinguish the two last from *sz* and *zē* followed by *hē*, we may either mark the *h*, or, as is common underline the digraphs, which is, however, more troublesome for the printers. On this account, *sh* and *zh* might be preferable to *sh̄*, *zh̄*, &c.

For the very deep guttural *qāf*, as in „Qorān“, the letter *q*, not otherwise required, is the best of all representatives: to use *k*, with the dot appropriated to the linguals, is altogether unjustifiable on any grounds.

There is hardly an alternative to *gh* and *kh* as representing *ghāin* (غ) and *khā* (خ); and the soft aspirate (ع) has been so generally represented by the usual apostrophe, that it may be left, and the reversed form used for the *small 'aïn* or *hamza*.

The vowels in Turkish are so variable that they can be indicated only very generally and imperfectly. The general system of transliteration for the four tongues is summarized in the second table.

SECOND TABLE.

THE ARABIC, TURKISH, PERSIAN, & HINDUSTANI
ALPHABETS.

CLASS	LETTERS	NAMES	ARABIC	TURKISH	PERSIAN	HINDUSTANI	
Vowels	ا	<i>Fatha</i>	a	a	a	a	
	آ	<i>Madda</i>	ā	ā	ā	ā	
	إ	<i>Kasra</i>	i	i	i	i	
	ـِ	<i>Zamma</i>	—		i	i	
	ي		ī	ī	ī	ī	
	و		u	u	u	u	
	ـُ		ū	ū	ū	ū	
	يَ	<i>Hamza</i>	—	e	e	e	
	أَ		ai	ai	ai	ai	
	و	<i>'Ain</i>	—	o	o	o	
	أَ		au	au	au	au	
	Gutturals & Fauicals	ء	<i>Hamza</i>	‘	—	‘	‘
		ع	<i>'Ain</i>	‘	‘	‘	‘
		ح	<i>Hē</i>	h	h	h	h
خ		<i>Hā</i>	ḥ	ḥ	ḥ	ḥ	
ق		<i>Qāf</i>	q	q	q	q	
ك		<i>Kāf</i>	k	k	k	k	
ج		<i>Gāf</i>	—	g	g	g	
ج		<i>Ṣaghir-nūn</i>	—	ñ	—	—	
ك		<i>Khā</i>	kh	kh	kh	kh	
ج		<i>Ghāin</i>	gh	gh	gh	gh	
Palatals	چ	<i>Chē</i>	—	ch	ch	ch	
	ج	<i>Jim</i>	j	j	j	j	
	ش	<i>Shīn</i>	sh	sh	sh	sh	
	ژ	<i>Zhē</i>	—	zh	zh	zh	
	ي	<i>Yē</i>	y	y	y	y	

THE ARABIC, TURKISH, PERSIAN, & HINDUSTANI
ALPHABETS.

CLASS	LETTERS	NAMES	ARABIC	TURKISH	PERSIAN	HINDUSTANI		
Dentals	ت د ن س ل ر ث ز	<i>Tē</i>	t	t	t	t		
		<i>Dāl</i>	d	d	d	d		
		<i>Nūn</i>	n	n	n	n		
		<i>Sīn</i>	s	s	s	s		
		<i>Lām</i>	l	l	l	l		
		<i>Rē</i>	r	r	r	r		
		<i>Sē</i> or <i>Thē</i> *	ṭh	s	s	s		
		<i>Zāl</i> or <i>Dhāl</i> *	dh	z	z	z		
		<i>Zē</i>	z	z	z	z		
		Linguals & Cerebrals	ا ت ت ت ت ت ت ت ت ت ت	<i>Ṭa</i>	—	—	—	ṭ
				<i>Ḍa</i>	—	—	—	ḍ
				<i>Ra</i>	—	—	—	r
<i>Ṭā</i> †	ṭ			ṭ	ṭ	ṭ		
<i>Ṣād</i>	s			s	s	s		
<i>Zād</i> or <i>Dād</i> *	ḍ			z	z	z		
<i>Zā</i> †	z			z	z	z		
Labials	ب ف م و			<i>Pē</i>	—	p	p	p
				<i>Bē</i> or <i>Bā</i>	b	b	b	b
				<i>Mīm</i>	m	m	m	m
		<i>Fē</i> or <i>Fā</i>	f	f	f	f		
		<i>Wāw</i>	w	w, v	v	v		
+Or	ط ظ	<i>Ṭa</i>	ṭ	ṭ	ṭ	ṭ		
		<i>Zā</i>	z	z	z	z		

* Three letters lisped in Arabic.

In the preceding remarks I have attempted to apply the scientific principles, so admirably expounded by professor Lepsius, to the transliteration of the Arabic and Persian alphabet, with as little disturbance to systems hitherto employed as practicable. Where several ways of representing an Oriental letter have been used, it was necessary to select one, — not arbitrarily, — but, first, in accordance with the classification of sounds, and, secondly, — and so far as consistent with this, — one that has been largely used. No new forms have been proposed; and the convenience of the practical printer has been considered. Where there is divergence of representation among scholars, whether in Sanskrit or Arabic, especially where the letters, without diacritical marks, do not convey to the reading public a fair idea of the sounds intended, — as in the use of *c* and *s* to represent a palatal and cerebral respectively, do I propose that practical considerations should prevail. Uniformity in the system of transcription would be worth much larger sacrifices than anything above suggested, and there is no hope of attaining that if we refuse to yield some small points for sake of it.
