

# PROBLEMS AND PITFALLS IN THE PHONETIC INTERPRETATION OF KHASI ORTHOGRAPHY

†Eugénie J. A. Henderson

The growing interest in Khasi in recent years, as the lone representative of the Austroasiatic family in an area surrounded by Indo-Aryan and Tibeto-Burman languages, has resulted in a welcome number of serious studies of the language. A number of otherwise valuable studies have, however, been flawed by misunderstanding arising from the Khasi spelling system. This looks relatively simple on the surface but there are pitfalls for the unwary. Scholars with an English or Indo-Aryan linguistic background may be led astray if they are tempted to take the spelling at its face value and to trust their eyes rather than their ears. The sections devoted to phonetics and phonology in some studies I have seen have contained phonetic or phonemic 'transcriptions' which are little more than letter-for-letter replicas of the spelling.<sup>1</sup> In these circumstances, it is worthwhile taking a closer look at problem areas within the spelling system itself.

There have, I believe, been certain attempts during the last few years at minor revisions of the official orthography, and I cannot claim to have up-to-date knowledge of all of these. As far as I am aware, however, the system is substantially still the same as that used in Diengdoh (1965), Kharkongngor (1968), Bars (1973), and Blah (1974). I shall also refer in this paper to the older and probably better-known dictionary of U Nissor Singh (1906), and occasionally to even older forms. The pronunciations cited are from Cherrapunji speakers, since this variety of Khasi is accepted as 'standard'.

## *Vowel length and final consonants*

What has to be borne in mind is that the Khasi roman spelling was devised in the first half of the nineteenth century, not by English or Scots or American missionaries, as in some other parts of India and Burma, but by *Welsh* missionaries, who introduced certain Welsh orthographic conventions unfamiliar to most English speakers.

To the non-Welsh, one of the most confusing of these conventions relates to the use of the letters *-p*, *-b*, *-t*, *-d*, *-c*, *-j*<sup>2</sup> in word final position. Linguists accustomed to the phonemic differentiation of voiced and voiceless stops, in such languages as French and English for example, find it natural to suppose that Khasi has this sort of phonemic distinction also,

---

1. This does not apply to Rabel (1961), whose phonetic observations are beyond reproach.

2. On the absence of final *-g* in Khasi, and on the special distribution of final *-k*, see Henderson (1965).

when they are confronted for the first time with spellings such as those in (1) below:

- |           |                            |                                  |
|-----------|----------------------------|----------------------------------|
| 1. Khasi: | <i>dab</i> 'bullock'       | <i>dap</i> 'full'                |
|           | <i>kad</i> 'to tear'       | <i>kat</i> 'so much, as much as' |
|           | <i>mad</i> 'to taste, try' | <i>mat</i> 'joint'               |
|           | <i>sngab</i> 'gill'        | <i>sngap</i> 'to hear'           |

For those whose work has hitherto been mainly with Indo-European languages, there is nothing very remarkable about this kind of final alternation, but those who have specialised in the languages of South-East Asia to which Khasi is related, it would be not a little surprising to find a phonemic distinction between voiced and voiceless final stops. The characteristic typological pattern found all over continental South-East Asia is for the voice : voiceless contrast, which is frequently found in initial position, to be neutralised in final position. What we usually get is an unexploded voiceless stop, often accompanied by simultaneous glottal closure. Listening to the pronunciation of the Khasi words in question soon makes it clear that Khasi is no exception to the general rule—the final stops are voiceless, no matter what letter they are written with. What we also observe is that, in the great majority of cases, where the final letter in the spelling is *b* or *d* the preceding vowel is always long, while where it is *p* or *t* it is almost always short.<sup>3</sup> The words in (1) above are regularly pronounced:

- |    |                       |                      |
|----|-----------------------|----------------------|
| 2. | <i>dab</i> [da:p̚]    | <i>dap</i> [dap̚]    |
|    | <i>kad</i> [ka:t̚]    | <i>kat</i> [kat̚]    |
|    | <i>mad</i> [ma:t̚]    | <i>mat</i> [mat̚]    |
|    | <i>sngab</i> [sɲa:p̚] | <i>sngap</i> [sɲap̚] |

This seems very curious until it is realised that Welsh has just such a convention: vowels before the letters *b*, *d* and *g* are almost always long, whereas before *p*, *t* and *c* are almost always short. (For this rule, and for the exceptions to it, see Wells (1979).)

- |           |                    |            |                   |               |
|-----------|--------------------|------------|-------------------|---------------|
| 3. Welsh: | <i>mab</i> [ma:b]  | 'son'      | <i>map</i> [map]  | 'map'         |
|           | <i>tad</i> [ta:d]  | 'father'   | <i>at</i> [at]    | 'to, towards' |
|           | <i>llog</i> [lɔ:g] | 'interest' | <i>llac</i> [lak] | 'slack'       |

Phoneticians who have worked on Welsh claim that while the sounds written *-b*, *-d*, *-g* are voiceless finally, they are also lenis, or lax, as compared with those written *-p*, *-t*, *-c*, which are voiceless and fortis, or tense. Wells thus prefers to describe the lenis set as 'devoiced', here symbolised [b̥, d̥, g̥], following the IPA convention. In Khasi there is no such tense/lax distinction, any more than there is a voice/voiceless distinction in final stops: they are all voiceless. It seems clear that the early Welsh missionaries made use of the their own spelling conventions in

---

3. See note 2 above.

order to mark distinctions of vowel length in Khasi.

This convention works well with stops, where there are two sets of letters available, but it does not work for words ending in final nasals or in final *r*, which are always voiced. In these circumstances, in Welsh, a circumflex accent is used to mark the long vowel:

4. Welsh:	<i>tân</i> [ta:n]	'fire'	<i>tan</i> [tan]	'under'
	<i>tôn</i> [to:n]	'tune'	<i>ton</i> [ton]	'wave'
	<i>glân</i> [gla:n]	'clean'	<i>glan</i> [glan]	'bank'

Earlier spellings of Khasi also tried this device. In Nissor Singh (1906) the circumflex accent is often used to mark vowel length before nasals and *r*; alternatively, a grave accent may be used to show length, or an acute accent to show shortness, but this is not at all regular, e.g.:

5. Khasi:	<i>bâm</i> [ba:m]	'to eat'	<i>sbâi</i> [sba:i]	'cowry'
	<i>kêr</i> [kɛ:r]	'to enclose'	<i>kpêr</i> [kpɛ:r]	'garden'
	<i>hêr</i> [hɛ:r]	'to fly'		
	<i>ûr</i> [ʊr]	'to slip, fall'	<i>tur</i> [tʊr]	'to butt'

It appears probable that, in these words, there is some variation in usage among Khasi speakers, which may account for some of the seeming irregularities. Thus, we find *ûr* 'to slip, fall', marked with a short vowel in Nissor Singh, but *tur* 'to butt' written without an accent, although also pronounced with a short vowel by my Khasi informants. Similarly, we find the spelling *kêr* [kɛ:r] 'to enclose', but *kpêr* [kpɛ:r] for the derived form meaning 'garden, enclosure'. *Hêr*, marked short, was pronounced with a long vowel by my informants, but it may be that some speakers use a short vowel here. This irregularity and uncertainty in the use of accents may be one of the reasons why they seem to have been abandoned completely in most of the recent dictionaries. This leaves the problem, for the foreigner, of being unable to tell from the spelling whether the vowel is long or short in such words. Father Bars' (1973) dictionary, for example, has to resort to such entries as '*bam* (long a)' and '*bam* (short a)'.

#### *The vowel 'y'*

Another problem for linguists has been the vowel spelt *y* in Khasi. It is almost always unstressed and behaves in much the same way at the *shwa* vowel [ə] in English, as in 'about, again, collapse, suppose', etc. When we look at Welsh, we find this similarity is no accident, since *y* is the Welsh way of spelling the vowel [ə]. Depending upon its position in the word, *y* may in Welsh also represent a vowel which is pronounced either [ɪ] or [i], according to dialect, e.g.: *mynydd* ['mənɪð, 'mənɪð] 'mountain'. *y*, with the [ə] value, is also the vowel that is used where a cushion vowel or epenthetic vowel is felt to be needed to break up awkward or alien consonant clusters, as in the forms *ysgôr* ~ *sgôr* [əsgo:r, sgo:r] 'score'. The rather old-fashioned Khasi spellings *kypa* 'father', *bydi* 'twenty', *bynâi* 'moon', month', etc., show the same process at work. Since clusters like [kp, bd,

bn], and many others, are difficult for Welsh or English speakers to pronounce without a short vowel in between, the *y* vowel was inserted. More recently the convention on the whole is not to write *y* between 2-consonant clusters, but to insert it where there are clusters of more than two consonants, e.g. *bnā* 'to hear', *bdi* 'twenty', but *byndi* 'to mortgage'. There seems still to be some uncertainty about this, however, as Father Bars still gives spellings like *byna*. Phonologically speaking, there would be much to be said for dropping the *y* in the 3-consonant clusters as well, simply writing *bndi*, which is perfectly clear. This is the solution proposed by Dr. Rabel in her phonemic analysis of Khasi (Rabel 1961).

### *Final palatals*

Khasi has final palatal consonants; Welsh does not. What appears to have struck the ears of the missionaries who first listened to Khasi was the *i*-like glide that precedes the final palatals. We thus have early spellings such as *skáin* 'fly' *n.*, *bysein* 'snake, python', etc., and it seems clear that the final consonant was thought of as plain *n* in such words, though its palatal quality was later recognised by the use of the letter *ñ*: *skaiñ*, *bseiñ*, etc. Since there is no accepted roman equivalent for the final palatal stop, *-d* and *-t* continued to be used, except after *ie*, when we find *j*, e.g.:

6. Khasi	<i>skaiñ</i>	[ska:j]	'fly n.'	<i>kruiñ</i>	[krɔj]	'white ant'
	<i>kshaid</i>	[kʃa:cʰ]	'waterfall'	<i>kait</i>	[kacʰ]	'plantain'
	<i>btuid</i>	[btu:əc]	'slippery'	<i>buit</i>	[bucʰ]	'cunning'
	<i>miej</i>	[mi:əc]	'table'			

From the phonological point of view, the vocalic segments preceding these consonants should be interpreted as plain vowels (or in some cases as centring diphthongs), with an automatic palatal on-glide to the final consonant, rather than as phonemes /ei, ai ui/, etc. Father Schmidt, writing on Khasi as early as 1904, recognised this; his knowledge of the Austroasiatic family led him to expect that final palatals might be present and he interpreted the missionaries' spelling in this light.

### *Representation of the glottal stop*

The representation of the glottal stop, which is a consonant phoneme in Khasi, does not derive from a Welsh convention, but rather from a much more general missionary usage, found also in the roman orthographies of Tibeto-Burman languages such as Lushai and Chin. Word-initially, the glottal stop is not written at all; word-finally it is spelt with *-h*; which is not otherwise needed in final position:

7. Khasi	<i>khlieh</i>	[khleʔ]	'head'	<i>spah</i>	[spaʔ]	'hundred'
	<i>soh</i>	[səʔ]	'fruit'	<i>kseh</i>	[kseʔ]	'pine'

Problems arise when [ʔ] occurs medially, or when it is the second consonant of an initial cluster. Devices used in such contexts include hyphens, apostrophes, or the vowel letter *y*. In the last case, the *y*

presumably represents the slight vocalic sound that can be heard as the glottal stop is released:

8. Khasi	<i>ha-oid, ho-oid, h'oid</i>	[haʔɔ:cʔ]	'Yes';
	<i>sh'iap, shyiap</i>	[ʃʔiapʔ]	'sand';
	<i>l'er, lyer</i>	[lʔɛ:r]	'wind';
	<i>s'iar, syiar</i>	[sʔiar]	'fowl';
	<i>lyoh</i>	[lʔɔʔ]	'cloud';
	<i>syiem</i>	[sʔi:əm, sʔe:m]	'king, chief'

In the word for 'cloud' above, we find the glottal stop represented by *h* finally, and by *y* in the initial cluster.

#### *Dental/alveolar distinction*

It is common for Western linguists to expect that *t*, *th*, *d*, *n* will form a natural class, united by a common place of articulation. In Khasi, however, as in certain other languages in the area, this turns out not to be the case. [t] and [th] are dental, [d] and [n] are alveolar. This is not simply a matter of phonetic detail; it is important in understanding the rules for permitted sequences of plosives and nasals in Khasi initial clusters. At first sight it appears that a sequence of virtually any two consonants may form an initial cluster, but while the possibilities are very rich, there is a constraint upon sequences of plosives and nasals whereby those with the same place of articulation may not form a cluster together. For example, *bn*, *bt*, *bth*, *pd*, *pn*, *bn*, *thɲ*, *dɲ*, *dp*, *km*, *kt*, etc., are permitted; *pb*, *pm*, *bm*, *kɲ*, *khɲ*, are not; *td*, *tn* and *thn* are permitted, and seem to be exceptions until it is recognised that the sequences here are also hetero-organic, and hence perfectly regular.

#### *Ambiguous vowel spellings*

The letters *u*, *o*, and *e*,<sup>4</sup> and the digraph *ie* are sometimes phonetically ambiguous.

Finally and before a glottal stop, *ie* represents a pure vowel, a very close [ɛ]. Before other consonants, pronunciation varies between [i:ə] and [e:], depending upon the speaker:

9.	<i>khlieh</i>	[khleʔ]	'head'	<i>lieh</i>	[leʔ]	'white'
	<i>sdiē</i>	[sdeʔ]	'axe'	<i>kmie</i>	[kmeʔ]	'mother'
	<i>dieng</i>	[di:əŋ, de:ŋ]	'tree, wood'	<i>ktien</i>	[kti:ən, kte:n]	'word'

In open syllables, the spelling *o* represents a very close [o] in some words, [ɔ] in others. There seems to be variation in usage between speakers for some words:

10.	<i>ro</i>	[ro· or rɔ·]	'quicksilver';	<i>to</i>	[tɔ] (indicating assent);
	<i>-to</i>	[tɔ·]	'that (enclitic)'.		

In closed syllables, the phonetic value is always [ɔ].

---

4. For a more detailed account of Khasi vowels, see Henderson (1967).

*u* before a final glottal stop is [uː] in some words, [oː] in others:

- |     |            |        |                |               |         |        |
|-----|------------|--------|----------------|---------------|---------|--------|
| 11. | <i>ruh</i> | [ruːʔ] | 'also, indeed' | <i>ruh</i>    | [roːʔ]  | 'cage' |
|     | <i>duh</i> | [doːʔ] | 'lose'         | <i>shniuh</i> | [ʃnoːʔ] | 'hair' |

Before other consonants, *u* may represent long [uːə]. [oː] or short [u]:

- |     |            |                 |          |             |                   |                   |
|-----|------------|-----------------|----------|-------------|-------------------|-------------------|
| 12. | <i>bud</i> | [buːətː, boːtː] | 'follow' | <i>thub</i> | [thuːəpː, thoːpː] | 'reduced in size' |
|     | <i>tup</i> | [tɒpː]          | 'cannon' | <i>lüm</i>  | [luːəmə, loːmə]   | 'hill'            |
|     |            |                 |          | <i>lüm</i>  | [ləmə]            | 'gather'          |

In open syllables, *e* may represent either close [e] or open [ɛ]:

- |     |           |       |             |           |       |         |
|-----|-----------|-------|-------------|-----------|-------|---------|
| 13. | <i>te</i> | [teː] | 'then, but' | <i>de</i> | [dɛː] | 'also', |
|     | <i>me</i> | [mɛː] | 'thou'      | <i>re</i> | [reː] | 'or'    |

In closed syllables the value is always [ɛ].

To sum up, with a proper understanding of its origins and background, Khasi spelling has much to recommend it. Once mastered, the conventions regarding the use of the letter *y*, the use of *-b*, *-d*, *-j* to indicate length of the preceding vowel, and the conventions for the representation of the glottal stop, are in the main clear and unambiguous. A reasonable complaint might be that vowel length before nasals is commonly *not* shown. However, in my view, the traditional spelling, properly understood, represents Khasi usage far more accurately than the misinterpretations and attempted 'improvements' that have been suggested by some linguists with an Indo-Aryan linguistic background, and little or no understanding of the strong Mon-Khmer legacy that still survives in Khasi phonological structure.

# REFERENCES

- Bars, E. 1973. *Khasi-English dictionary*, Shillong: Don Bosco Press.
- Blah, V.E. 1974. (comp.) *Chapala's Anglo-Khasi dictionary*. (3rd ed.) Shillong: Chapala Book Stall.
- Diengdoh, A.K. 1965. *Leemuel's Anglo-Khasi pocket dictionary*. Shillong: L. Harrison.
- Henderson, Eugénie, J. A. 1965. Final -k in Khasi: a secondary phonological pattern. *Lingua* 14, 459-66.
- ..... 1967. Vowel length and vowel quality in Khasi. *Bull. Sch. Or. Afr. Stud.* 30(3), 564-88.
- Kharkongngor, U I. 1968. *Ka Dienshonhi (a Khasi-Khasi dictionary)* Shillong: Ri Khasi Press.
- Rabel, Lili 1961. *Khasi: a language of Assam*. Baton Rouge: Louisiana State Univ. Press.
- Schmidt, W. 1904. *Grundzüge einer Lautlehre der Khasi-Sprache in ihren Beziehungen zu derjenigen der Mon-Khmer Sprachen*. München: Verlag der K. Akademie.
- Singh, U Nissor 1906. *Khasi-English dictionary*. Shillong: Eastern Bengal & Assam Secretariat Press.
- Wells, J.C. 1979. Final voicing and vowel length in Welsh. *Phonetica* 36(405), 344-60.