

KHASI INITIAL CLUSTERS

Eugénie J.A. Henderson

School of Oriental and African Studies,
University of London

1.0. One of the most striking features of Khasi is the astonishing richness of its word-initial clustering possibilities. This great variety of initial clustering patterns is a typological feature of other Austroasiatic languages, as for example Khmer and Old Mon, and contrasts sharply with neighbouring Sino-Tibetan languages in which permitted initial clusters are relatively few in number and restricted almost exclusively to plosives plus following liquids or semivowels (l, r, y, w). Rabel, who describes the variety of Khasi initial clusters as "absolutely phantastic" (Rabel 1973), lists 127 two-consonant clusters with examples of the occurrences of each with each of twelve final consonants (Rabel 1961, 21-29). Taking her careful study as a starting point I have even been able to add a few more to this list though it may turn out that some of the newer "discoveries", *e.g.* sr, may be stylistic variants of others already listed (see 5.12.).

1.1. Pater Schmidt's theory was that the Mon-Khmer syllabic pattern for what he called 'stems' was CVC; hence all words of the structure CCVC etc. were interpreted by him as being "stems" with either prefixes or infixes added (Schmidt 1904). This seems to be an extreme view. There is surely no reason to suppose that Mon-Khmer languages could not have some

ial clusters which are *not* dimorphemic, but
idit was able to support his thesis in many in-
ces by references to non-affixed stems in cognate
guages (see Henderson 1973). As Greenberg has
n, morphological considerations may have an im-
tant bearing upon the clustering patterns of a
guage (Greenberg 1965), especially as regards ap-
ent "exceptions." It behooves us to bear such
siderations in mind in the investigation of Khasi
cial clusters, which are interesting not only from
comparative/historical point of view but also for
ir relevance to the current pursuit of "language
versals."

. This paper will be concerned only with the
ly of two-consonant initial clusters, which are
more numerous and far less restricted in their
binations than initial groups of three or more
sonants. The permitted combinations of two conso-
ts seem at first sight so generous and so varied
t one is inclined to believe that virtually *any*
bination is possible. Further probing, however,
eals that there are certain gaps and that there
also certain restrictions upon the occurrence of
permitted combinations. It is proposed first to
t the permitted types of clusters, then to draw
ention to certain restrictions on the occurrence
some of these, and lastly to examine the "gaps"
h a view to determining how far these may be
arded as systematic rather than accidental.

Orthography, Glosses and Sources.

. *Orthography.* Khasi examples are cited in the
epted roman orthography, with the minor modifica-
ns set out in Henderson 1973, Section 1.1. Re-
ders of IPA values are sometimes added in square

brackets. It should further be noted that final -j represents a checked voiceless palatal stop [c]. Words such as riat [riat] 'precipice' might be interpreted as comprising an initial cluster [rj-] + [at]. Such an interpretation would, however, lead to three-consonant clusters such as [ksj-] in ksiaɾ 'gold', [bsj-] in bsiang 'pitch' etc. which are "irregular" when compared with the vast majority of three-consonant initial groups, in which, in true Mon-Khmer fashion, the middle consonant is regularly a liquid or a nasal (see Rabel 1961; Henderson 1973). I have therefore decided here to regard ia as representing a vocalic element, and have not posited any initial clusters containing the palatal semivowel [j].

2.2. *Glosses.* The cautionary note about the English glosses of Khasi words contained in Henderson 1973, Section 1.2., also applies here.

2.3. *Sources.* The data were obtained in the course of work with Namita Shadap Sen and Ivan Simon (see Henderson 1973), and compared with Rabel 1961 and Nissor Singh 1906. The language under discussion is standard Khasi as used by educated speakers in Cherrapunji.

3.0. *Permitted Clusters.*

The types of two-consonant cluster permitted at the beginning of Khasi words are listed below with examples of subtypes within each.

The glottal stop has for convenience been grouped with the unaspirated voiceless stops.

w-clusters have been listed together under one heading at 3.33, since w is the only semivowel found as a component of clusters (but see 2.1 above).

j-clusters are also listed separately at 3.31 3.32 since I am undecided whether to classify si j with the voiced plosives b and d, or with the bilants' s and sh.

The 'voiced aspirates' bh and jh also pose problems. jh can be justified as a cluster by virtue of the shortened combining form 'hur of jhur 'vegetables'. I know of no similar short combining form in any word beginning with bh however, and it is doubtful whether bh should be included among the clusters here. The unexpected presence of jh and bh in the Mon-Khmer language immediately suggests Indo-European borrowing and this is indeed clearly the origin of these sounds in a number of instances. Whether it can account for *all* instances is perhaps not so certain, despite the persuasive arguments of Rabel (Rabel 1973). That Khasis should have borrowed such a common everyday word as bha 'good, well' from some neighbouring IA language is to me surprising, and even considering Rabel's suggested derivation (Rabel 1973) from such forms as Bengali bhalo, Assamese bhal, one must also take into account the Northern Khasi (Mnar) dialect form bhi. And what of the name bhi for an inhabitant of the northern lowlands? An Assamese word meaning 'fear' is not a wholly convincing origin. It does not seem beyond the bounds of possibility that there may be a merging of forms from indigenous as well as foreign sources in words containing these sounds.

1. Voiced stop + voiced stop:

bdi 'twenty'

2. Voiced stop + aspirated voiceless stop:

bthi 'greasy', bthong 'hog-deer', bthuh 'grey', bthei 'to explode'.

dkhar 'plainsman, non-Khasi', dkhiew 'ant',
dkhoh 'owl', dkhap 'tick', dkhat 'to
break off, snap', dkhi 'wasp'.

3.3. Voiced stop + unaspirated voiceless stop:

bta 'to wash the face', btái 'tired', bti
'poma tree', bteng 'to join', btuid 'to
slip'.

dkár 'tortoise', dkái 'honey-fly', dkoh
'lame'.

dpei 'ashes'.

3.4. Unaspirated voiceless stop + voiced stop:

pda 'to fling', pdot 'throat', pdeng 'mid-
dle', pdem 'to steep'.

kba 'paddy', kber 'winged white ant', kboh
'to scratch', kdang 'to be seated',
kdait 'reed', kdew 'to point', kdong
'corner'.

tba 'to explore with the fingers', tbit
'skillful', tbeh 'to light', tbian
'floor', tda 'span', tdem 'smoke',
tdong 'tail'.

3.5. Unaspirated voiceless stop + unaspirated voiceless stop:

ptej-ptej 'glutinous', pylar [pʔ-] 'to
stretch'.

tkai-tkai 'plump', tker 'pierced and quiver-
ing', tyut [tʔ-] 'quail'.

kpa 'father', kper 'garden', kpu 'loaf',
kpoh 'abdomen', kti 'hand', ktieh 'mud',
ktang 'water pipe', kya [kʔ-] 'simul
tree', kying [kʔ-] 'hornet'.

3.6. Unaspirated voiceless stop + aspirated voiceless stop:

ktha 'to chew', kthang 'bitter', ktháw
'father-in-law'.

3.7. Voiced stop + nasal:

bna 'to hear', bnái 'moon', bniat 'tooth',
bñiup-bñiup 'drizzling'.

dnglem 'bear', dngiang 'lively'.

9. Aspirated voiceless stop + nasal:

phna 'to fasten (a door)', phniang 'seed',
phñian 'to stow', phngáin 'clear (of
sky)', phngóit 'to peep out'.

thma 'war', thmu 'to resolve', thning
'stalk', thngan 'hungry'.

khmat 'face, eye', khmut 'noise', khnái
'mouse', khnam 'arrow', khñi 'to whim-
per', khñiang 'insect'.

10. Unaspirated voiceless stop + nasal:

pnar 'shell', pnat 'forked branch'.

tnga 'spouse', tnat 'twig', tnúm 'house,
ridge', tmaiñ 'moustache', tmoh 'chin',
tmier 'rim'.

kmie 'mother', kmen 'to rejoice', knah
'scaffold', kñi 'maternal uncle', kñia
'to sacrifice'.

11. Voiced stop + liquid:

blei 'god', blang 'goat', bri 'grove', briew
'man, person', brot 'bronze'.

12. Aspirated voiceless stop + liquid:

phlang 'grass', phlei 'to vomit, flood',
phra 'eight', phrung 'to insert'.

thliem 'leech', thlong 'mortar', thri 'bam-
boo tie', thrang 'to pant, thirst'.

khla 'tiger', khlieh 'head', khluér 'star',
khra 'pitcher', khriat 'cold'.

13. Unaspirated voiceless stop + liquid:

pla 'bag', pliang 'plate', plied 'to unfold',
prek 'nail', pring 'soot', prúm 'ulcer'.

tlang 'winter', tlái 'palmtree', trái
'root', trei 'to work', treng 'mane'.

klew 'peacock', klong 'heart, gourd', kren
'to speak', kriah 'sieve', krong 'bier'.

14. Voiced stop + sibilant:

bsa 'to feed', bseiñ 'serpent', bsong
'ferret', bshad 'civet'.

3.14. Voiceless stop + sibilant:

ksang 'bile', ksew 'dog', kseh 'pine tree',
kshaid 'waterfall', kshong 'pork fat'.

3.15. Voiced stop + aspirate (voiced aspirated stops)

bha 'good', bhah 'share', Bhoi 'inhabitant
of northern lowlands'. (See also 3.31
for jh-.)

3.16. Nasal + stop:

mtung 'with a large swelling', mthin 'stout'
mtai 'in a large heap', mdong '(an
onomatope)'.

3.17. Nasal + liquid:

mrad 'wild beast', mraw 'slave', mlien 'be
in the habit of', mluh 'salt'.

3.18. Nasal + nasal:

mnong, mnung 'sitting motionless by oneself'

3.19. Liquid + voiced stop:

rben 'thick', rdong '(held) securely'.
lbong 'thigh', Lber 'the month of March',
ldai 'dangling'.

3.20. Liquid + voiceless aspirated stop:

rkhle 'laugh', lkhai 'tender (of meat)',
lphong 'straight (like a post or tree)'

3.21. Liquid + voiceless unaspirated stop:

rta 'age', rye [rʔ-] 'full, bloated (of
stomach)', ryeng [rʔ-] 'propped up at
one side'.
lpa 'already', lko 'disintegrate', lyer
[lʔ-] 'wind, air', lyeit [lʔ-] 'but-
tocks', ltar 'at full length'.

3.22. Liquid + nasal:

rml 'astringent', rmiang 'margin, rim',
rnong 'brass', rniang 'anvil', rnga
'charcoal', rngái 'dream, illusion',
rñai 'a species of bamboo'.
lmun 'tolerant of unpleasant things', lmit-
lmit 'yellowish', lngaid 'fritter away
one's time', lngiar 'be late'.

23. Liquid + sibilant:

rsham 'sacrificial articles', lshulñ 'oval'.

24. Liquid + aspirate:

rhah 'envious', rhem 'glowing'.

lhen, lhon 'stagnant', lhop 'sultry'.

N.B.: Initial liquid + aspirate clusters are pronounced as a sequence of voiced liquid + (potentially voiced) aspirate, not as voiceless liquids.

25. Liquid + liquid:

rlang 'species of white ant'.

26. Sibilant + voiced stop:

sbái 'shell, money', sboh 'manure', sbur
'rich (of soil)', sdad 'confluence',
sdang 'begin', sdle 'axe'.

shdei 'showing the teeth'.

27. Sibilant + voiceless aspirated stop:

skhep 'hip', skhem 'firm'.

28. Sibilant + voiceless unaspirated stop:

spah 'wealth', sping 'handle', stang 'thin',
stem 'yellow', sting 'light (not
heavy)', skei 'barking deer', skap
'husk', skum 'nest', syiem [sʔ-]
'chief', sylar [sʔ-] 'fowl'.

shka 'junction of rivers', shkór 'ear',
shtah (or stah) 'to cut bark of a
tree', shyiap [ʃʔ-] 'sand', shying
[ʃʔ-] 'bone'.

29. Sibilant + nasal:

sma 'to stink', snad 'comb', snam 'blood',
snem 'year', sniew 'bad', sngew 'to
feel', sngi 'sun, day'.

shmoiñ '(full) to the brim', shnong 'vil-
lage, town', shniash 'chisel', shñluh
'hair'.

30. Sibilant + liquid:

sia 'leaf', slap 'rain', sriem 'similar'.

shlei 'to overflow', shlor 'brave', shreng
'fin', shrleh 'monkey', shrang
'parched'.

3.31. Clusters with j in first place:

jkek, jker 'pierced', jkeng 'tall and big',
jkhap 'with a snap'.

jnang 'to choke', jngái 'far', jngi 'to
swim'.

jlew 'to howl', jli 'hateful', jrong 'tall',
jmop-jmap 'a striking sound'.

jhur 'vegetables', jhieh 'wet', jhum
'vapour', jhap 'after birth'.

3.32. Clusters with j in second place:

kjat 'foot', pjah 'cold', kjup 'to suck'.

ljang 'hanging down'.

3.33. Clusters with w in second place:

khwái 'fishhook', kwah 'to wish for',
d(u)wai 'to pray'.

jwái 'to pinch', jwar 'birthmark', lwái
'wild honey', rwái 'to sing', rwat 'to
search', shwár 'an evil spirit', swai
'frail'.

thwlat 'to have a narrow escape', thwei
'deep pool', twa 'to subside'.

4.0. *Restrictions on Permitted Cluster Types.*

4.1. It is immediately noticeable that homorganic sequences are avoided: velars do not cluster with other velars, nor labials with labials, and so on. This Khasi characteristic will for brevity's sake be referred to as the 'heterorganic rule'. It should be noted here that td and tn, which both occur, are not homorganic, since t is dental while n and d are both alveolar.

There appears also to be a tendency to avoid labial-velar combinations, though, curiously, *not* to avoid velar-labial ones. kp- and kb- are quite

mon, whereas pk- and bk- do not occur. phng- is found in a few rare words, e.g. phngáĩñ 'to clear up (sky)', and certain phonaesthetic adverbs, used in stricted collocation, such as phngóĩt and phngong, which express the idea of appearing briefly or suddenly. p- and b- do not occur before any velars at 1.

2. Contrary to what Greenberg has put forward as a language "universal", or "near-universal" (Greenberg, 1965), Khasi has a number of stop clusters in which one element is voiced and the other voiceless. Not only does Khasi possess such clusters, the language appears to prefer stop clusters with voice dissimilation between the components to those in which there is voice assimilation between them; that is to say, t-, pd-, tb-, td-, kd- are the common, "preferred" forms whereas pt- and tk- are used only in phonaesthetic or "expressive" expressions, such as ptej-ptej, used to describe something sticky, glutinous; kái-tkai, describing someone plump; and tket or tker, used to describe a piercing movement (the final -r in the second form expresses vibration, so that tker could describe, for example, the piercing of an arrow, accompanied by quivering of the shaft).

There is one example only in Khasi of two voiced stops in an initial cluster, namely bd- in bdi, a rather rare word for 'twenty', used in counting firewood.

Khasi clusters such as bt-, dk-, bth-, dkh-, p- are further noted by Greenberg as an exception to the almost universal rule that in initial clusters in which there is one voiced component and one unvoiced component, the voiced component will be second, i.e.

nearest to the voiced vocalic nucleus of the syllable. So contrary to the general pattern are such Khasi clusters that Greenberg questioned whether they had in fact been correctly reported. Instrumental examination of Khasi clusters that I have undertaken confirms that the first element of a cluster such as *bt-* is indeed voiced, and that the second element is quite as clearly voiceless. There is, moreover, usually no perceptible vocalic glide between the two consonants as uttered by native Khasi speakers.

Clusters in which the first element is *k-* appear to be exceptions to the voice dissimilation rule, since *kp-*, *kt-*, *kth-* are at least as common as *kb-* and *kd-*. It is clearly relevant here that Khasi *k* has no voiced correlate *g-*,¹ hence the phonological status of *k* may be regarded in this context as having to do duty for both voiceless and voiced velar stops. There may also be morphological influences at work, since in many of the words having a *k-* initial cluster the *k* itself is the reflex of a former generic prefix for living things (see Henderson 1973).

4.3. Many of the permitted combinations are exclusively or almost exclusively reserved for families of "expressive" words with a phonaesthetic function, as for example:

ld- for dangling objects

lt- for objects stretched out full-length

mth- for big and muscular creatures

mt- for humped, swollen objects, heaps, etc

jn- descriptive of choking.

¹Except in a handful of loanwords. See short entry under this letter in Nissor Singh 1906.

er clusters which appear to be restricted to 'ex-
ssive' words are tk-, pt- (see 4.2), phng- (see
, md-, mn-, lph-, lm-, lsh-, shd-, jk-, jkh-, jm-
. (See also Henderson 1973, Section 10.)

. *Gaps.*

It is interesting to speculate how far the ap-
ent gaps among the Khasi clusters may be regarded
phonologically "systematic," and how far they are
result of historical "accident," the product per-
s of ancient morphological processes (see
nderson 1973). A full investigation of such "gaps"
are found in standard Khasi would of course re-
re an examination of the cluster patterns in other
si dialects also. What follows below is no more
an a listing of seeming "gaps," with such provi-
onal comments as occur to me in the present state
knowledge.

1. Aspirated stops are found only before liquids
l nasals, never before sibilants or other stops.

2. p- is not found before aspirated stops, sibi-
nts or before w. The absence before w is presu-
ly because of the restriction on homorganic
usters (see 4.1). The absence before aspirated
ops may be ascribed to the restriction on clusters
two voiceless stops (see 4.2), since bth- is found.
e absence before s or sh is more surprising, and it
noted that ps- clusters are found in certain Khasi
ialects, e.g. Mnar (N. Khasi) [psen] 'snake', War
amin dialect) [bɔ psɛu] 'brother-in-law'.

3. t is not found before sibilants, aspirated
ops, or ñ. The presence of t before all nasals
ner than ñ suggests, following the heterorganic
le, that Khasi t, which is dental, should be

aligned phonologically with ñ, sharing perhaps some such feature as [+laminal], which would imply articulation with the forward part of the tongue with the tip pointing downward, as contrasted with the apical but post-dental articulation of d and n. th, which is found before liquids and before all nasals except ñ, is a similar case. Possibly s, sh and j should be regarded as belonging to the same phonological set as t, th, and ñ.

5.4. As has already been pointed out (4.2) pt- and tk- are found in phonaesthetic words only. tp- and pk- might have been expected to occur in such words also, but in fact they do not occur at all. (For the absences of pk- see also 4.1.)

5.5. d is found before k, and kh (possibly as a 'dissimilated' version of the k- prefix, see Henderson 1973), before p (one instance only), and before ng and w. Before nasals other than ng, t and th are preferred. Perhaps the voice dissimilation rule applies here. There are no examples of clusters consisting of d + liquid, whereas thr-, thl-, tr-, and tl- are common. Since d is alveolar or post-alveolar, not dental like t and th, it may be that the heterorganic rule is operating here. Note also, however, the non-occurrence of d before s and sh.

5.6. The only nasal occurring as the first consonant in a cluster is m. Of such clusters, mt-, mth-, md- and mn- are restricted to phonaesthetic words. The only m-clusters not so restricted are mr- and ml-.

5.7. The only incidence of h as the first consonant of a cluster is in the word ha-oid (sometimes spelt h'oid) 'yes', which I interpret as [h?-].

. The glottal stop is never the first element of a cluster, but occurs as the second element in a number of clusters beginning with voiceless stops, sibilants and liquids, viz. [pʔ-, kʔ-, tʔ-, rʔ-, lʔ-, sʔ-, ...]. It does not occur after voiced stops, nasals or j.

9. The only "true" clusters (in the lexical sense) with j as the first consonant are jh-, jng-, jr-, and ... Other clusters with initial j appear to be restricted to words of the phonaesthetic type.

10. The absence of s before ñ, j or sh supports the suggestion that these consonants belong to the same phonological set (see 5.3). On the other hand, s occurs before t, which is also a consonant of this group.

11. sh is found before k, ng, n, r, l, d (in one or two phonaesthetic words), and the glottal stop. It is also found after b, k, r, l (phonaesthetic). This fits well with the suggested alignment of sh with the 'laminal' or 'tip down' set of consonants. There is, however, the common word shñiuh 'hair' which appears to break the heterorganic rule. Cognate forms in other languages show that the nasal here is an infix. Fixed nasals are frequently assimilated to neighbouring consonants and it may therefore be that in this instance the palatal character of the nasal is the result of assimilation to sh. Compare shñiuh 'hair' with shnīah 'chisel', in which the nasal is alveolar.

There is also the word shtah 'to cut the bark of a tree', which is in free variation with the form shah.

5.12. Only one example of sr- was found, namely sriem 'similar'. sr- here is possibly a stylistic variant of the relatively common cluster shr-. (See 5.11 for the variation sht ~ st.)

5.13. l occurs in "true" clusters before b, p, k, and the glottal stop. All other clusters with initial l are probably phonaesthetic. l does not occur before n, ñ, s, th, or r.

5.14. Initial r is found in the following clusters: rb-, rt-, rw-, rkh-, rm-, rn-, rng-, rd-, rsh-, rl-. The last cluster, consisting of two liquids, is a rarity in the languages of the world according to Greenberg (1965), who in the same paper claims that the reverse order, *i.e.* lr-, is never found (see 5.13). One might expect r to group with d and n in Khasi, but if so, rd- and rn- break the heterorganic rule. What then is the phonological alignment of Khasi r?

6.0. I do not feel able to close this paper without drawing attention to the curiously wasteful use (or rather lack of use) that Khasi makes of its great richness in clustering possibilities. Many clusters carry a very low functional load. There is only one word beginning with bd-, one with dp-, two with bsh-, and so on. To have so many permitted clusters and to make so little use of a great many of them strikes me as very puzzling. Moreover, there seems to be no inclination to make further use of the under-exploited clusters, except in the case of the phonaesthetic combinations.

REFERENCES

- eenberg, Joseph. 1965. Some generalisations concerning initial and final consonant sequences. *Linguistics*, 18.
- ndererson, Eugénie J.A. 1974. "Vestiges of Morphology in Modern Standard Khasi," in this volume.
- ssor Singh, U. 1906. *Khasi-English Dictionary*. Shillong: Eastern Bengal and Assam Secretariat Press.
- bel, Lili. 1961. *Khasi, a language of Assam*. Baton Rouge: Louisiana State University Press.
- bel, Lili. 1974. "Analysis of Loanwords in Khasi," in this volume.
- hmidt, Pater Wilhelm. 1904. *Grundzüge einer Lautlehre der Khasi-Sprache in ihren Beziehungen zu derjenigen der Mon-Khmer Sprachen*. Transactions of the Royal Bavarian Academy of Sciences, 22,3, 657-810.