

PHONOLOGICAL TYPOLOGY OF NORTHEAST INDIA*

Lukas Neukom
University of Zürich

1. INTRODUCTION

The area extending from Nepal in the north to Orissa in the south has been suggested by Ebert (1993) to have developed common features in some of the languages of the four distinct families represented there (Indo-Aryan, Dravidian, Austro-Asiatic and Tibeto-Burman). The aim of my paper is to present a phonological survey of the languages of an enlarged area, extending in the east to the border of Burma, thus including the Indian states of West Bengal and of Assam, the eastern states of India, Bangladesh and Bhutan as well.¹

In the domain of phonology some typologically marked features can be discovered in the four language families. For a first impression see Table 1 below, which compares the inventories of consonant phonemes of a representative of each family.

The similarity of the four inventories is obvious. Three of the phonological features to be examined in this paper are clearly visible in each language: a retroflex series of consonants, a breathy voiced plosive series, and a scarcity of fricatives.

The aim of my paper is to examine five phonological features that suggest themselves as characteristic for the area under investigation. These are the following:

1. retroflex phonemes
2. few fricatives

* This paper is a report on part of the project "Arealtypologie Südasien", supported by the Swiss National Science Foundation under Grant No. 1214-049271.96/1. The project deals with the convergence area of the south-western Tibeto-Burman languages with adjacent languages. I am indebted to Karen Ebert, Kathrin Cooper and John Peterson for helpful comments and suggestions. Errors are solely my own.

¹ According to Grierson (1903-28) the Tai language family as a fifth genetic unit is represented by two living languages (Aiton and Khamti) and one extinct language (Ahom) in India. I included Khamti in my survey, but apart from that I did not take the Tai languages into consideration.

3. breathy voice
4. creaky voice and related phenomena
5. initial velar nasal

The selection of the features is mainly based on lists proposed in the past, three of which (the first, third and fourth) are found, e.g., in Ramanujan and Masica (1969).² Masica (1976:187) quotes a further four features that have been postulated as areal features for South Asia: nasalized vowels; affricate opposition (**ts** vs. **tʃ**); syllabic structure and phoneme distribution; and tendency to initial stress. He denies or questions the validity of these four features. The list of features examined in this paper is confined to those that are rare enough, compared to the world's languages, to produce significant statements.

The languages included in the investigation are listed alphabetically in Appendix I, where they are presented along with bibliographic sources. I have tried to select at least one member of each sub-branch of the language families represented in the area, as shown in Appendix II. The classification is not very deep, especially in the case of Tibeto-Burman languages, where I mainly follow DeLancey (1987). This results in a survey over a range of 38 languages³ in the area in question, plus 26 languages of neighbouring areas, added by way of comparison. In reflecting on the historical development, only four of the six language families mentioned in the appendix are more closely examined, since the other two (Tai and Miao-Yao) play only a peripheral role in our area.

Two maps document the findings. Map 1 is a schematic representation of each of the languages of our area (and of the nearest neighbours) by squares, arranged roughly geographically. The squares contain the respective values for each feature, so that any exception is indicated. A key to the abbreviations of the languages is provided in the alphabetical list in Appendix I.

Map 2 situates our area in a slightly larger context and gives a somewhat more precise location for the languages. The map shows the attempt to plot correspondent isoglosses, neglecting minor deviations. Exceptions are marked by superposed small numbers, referring to the particular feature (e.g. ⁴ = language that shows a value for the feature of creaky voice contrary to what is implied by the isogloss).

² Other previous work on phonological typology relevant for our area includes Henderson (1965) for Southeast Asia and Michailovsky (1988) for Nepal. Both studies check an almost complete set of phonological features. Masica (1991:131) comments on the phonological properties of Indo-Aryan in comparison with the other languages of South Asia.

³ Including two dialects of Bengali (the dialect of Chittagong besides the colloquial standard) and two dialects of Kurux (Dhangar Kurux in Nepal, as well as the variant spoken in South Bihar).

1. **Indo-Aryan: Bengali**
(Ferguson & Chowdhury 1960)

p	t	ʈ	c	k
ph	th	ʈh	ch	kh
b	d	ɖ	j	g
bh	dh	ɖh	jh	gh
m	n			ŋ
			ʃ	
	l			
	r	ɽ		
[w]			[j]	

2. **Tibeto-Burman: Thulung**
(Allen 1975)

p	t	ʈ ⁴	c	k	(?) ⁵
ph	th	[ʈh]	ch	kh	
b	d	ɖ	j	g	
bh	dh	[ɖh]	jh	gh	
m	n			ŋ	
			s		h
	l				
	r	(ɽ)			
w			j		

3. **Austro-Asiatic: Kharia**
(Biligiri 1965)

p	t	ʈ	c	k	(?)
ph	th	ʈh	ch	kh	
b	d	ɖ	j	g	
bh	dh	ɖh	jh	gh	
(ʈb) ⁶	(ʈd)	(ʈj)			
m	n	(ŋ)	ɲ	ŋ	
	s				h
	l				
	r	ɽ (ɽh)			
w			j		

4. **Dravidian: Kurux**
(Pfeiffer 1972)

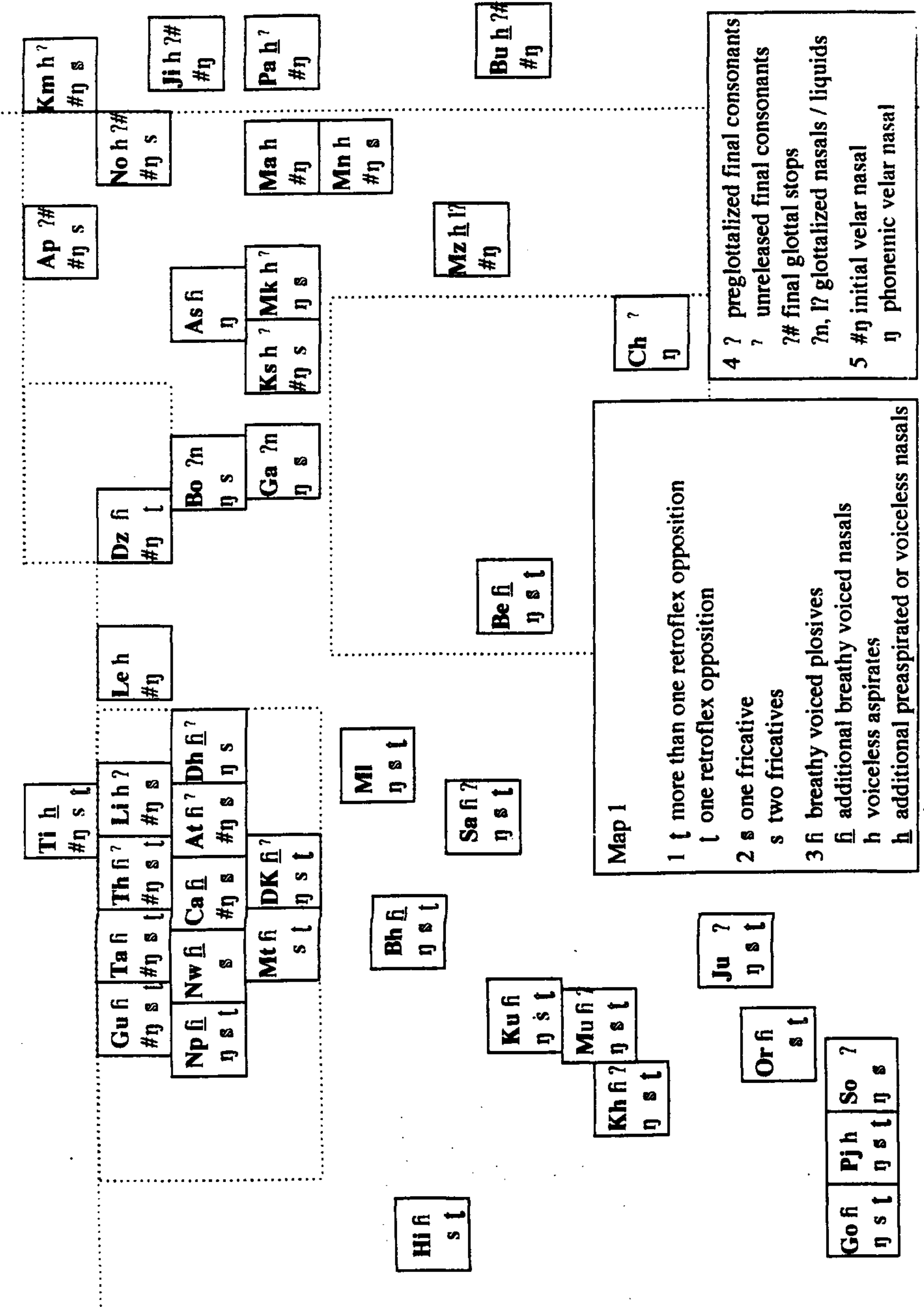
p	t	ʈ	c	k	?
ph	th	ʈh	ch	kh	
b	d	ɖ	j	g	
bh	dh	ɖh	jh	gh	
m	n	(ŋ)		ŋ	
	s			x	h
	l				
	r	ɽ (ɽh)			
w			j		

Table 1. Four consonant phoneme inventories

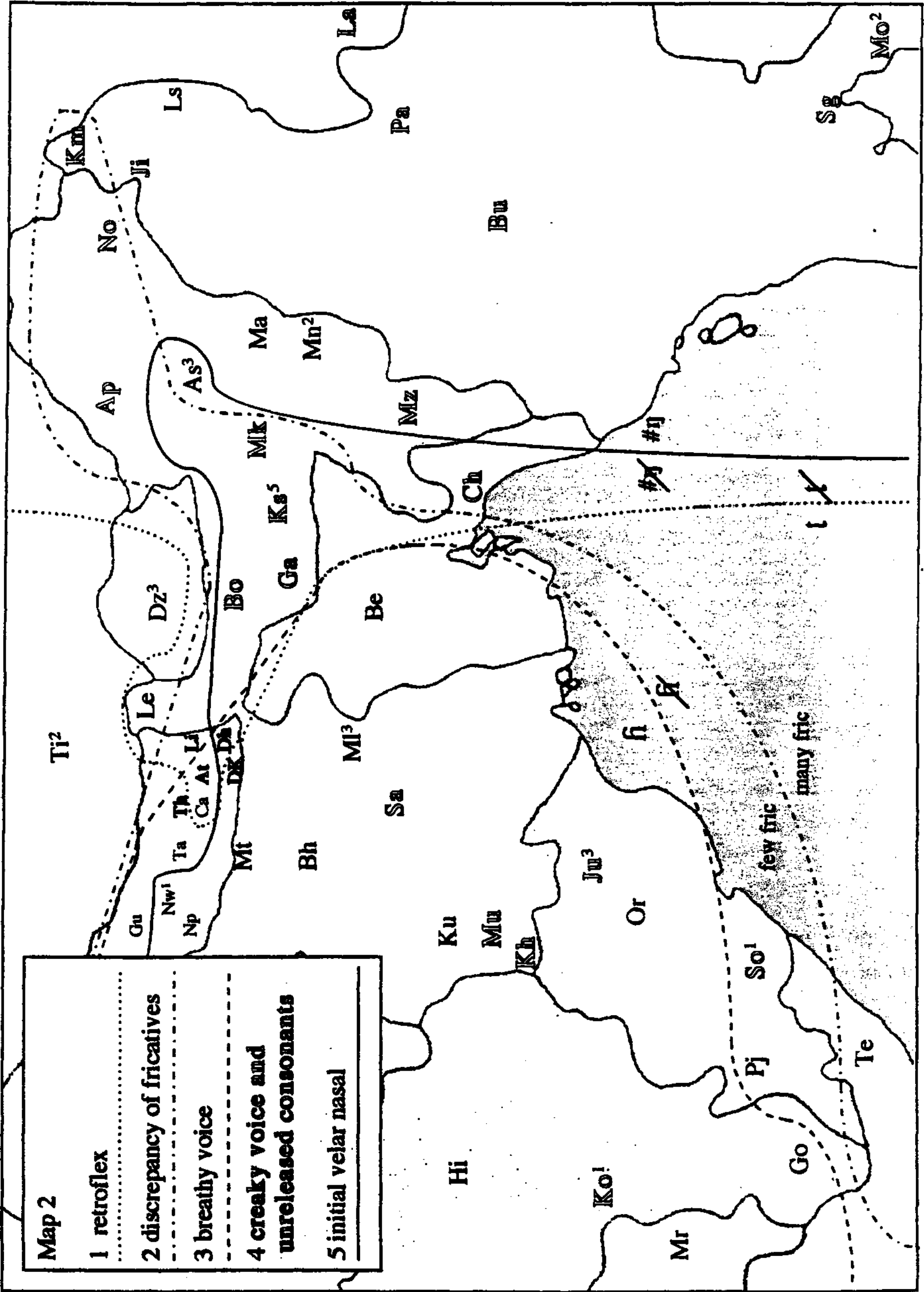
⁴ Retroflex phonemes are rather rare in Thulung. Allen (1975:14f) notes that aspirated retroflexes have been attested only in three words, and initially the voiceless retroflex ʈ occurs only in three items.

⁵ Symbols between parentheses are allophones, whereas symbols between square brackets have low functional load.

⁶ The status of the preglottalized consonants is controversial. They may be regarded as phonemes (Pinnow 1959).



Neukom, L. 1999, "Phonological typology of northeast India", in *Linguistics of the Tibeto-Burman Area*, vol. 22, no. 2, pp. 121-147. (purl.org/sealang/neukom1999phonological.pdf)



Neukom, L. 1999, "Phonological typology of northeast India", in *Linguistics of the Tibeto-Burman Area*, vol. 22, no. 2, pp. 121-147.
 (purl.org/sealang/neukom1999phonological.pdf)

I will now deal with each of the five features individually, their distribution, and their possible historical spreading. In the end a tentative overall picture will be given.

2. PHONOLOGICAL FEATURES

2.1. Retroflex

The term “retroflex” is used by many authors without explaining what exactly is meant (an exception is Mahapatra (1979:24) for Malto). It always denotes a curling of the tongue; the movement may be upwards and/or backwards, building a stricture at a position between alveolar and palatal.⁷ The important thing is that the sound called retroflex is clearly distinguishable from a dental consonant (Masica (1991:94): “in any case non-dental”). (Of course there are languages with a threefold contrast among dental, alveolar and retroflex, e.g., Malayalam, but no such language is found in our area.)

Retroflex sounds are not as rare as might be supposed. In Ruhlen’s (1976:157) sample of 693 languages of the world, 19% contain them. Their occurrence is concentrated in two areas: Australia (where 79% of his sample languages have retroflex sounds) and South Asia. The remaining languages containing them are scattered all over the world (for more details see Bhat 1973).

The retroflex isogloss, indicated by a dotted line, splits our area in two. Roughly speaking, the retroflex feature is distinctive in the western part, and absent or non-distinctive in the eastern part.

In the western part, south of the Nepal-India border, the dental-retroflex opposition is normally applied to more than one consonant type. While the contrast works in the plosive series (maximally **t** vs. **ɽ**, **th** vs. **ɽh**; **d** vs. **ɖ**, **dh** vs. **ɖh**) as usual, it is also present elsewhere, most often one of the liquids having a retroflex counterpart (**l** vs. **ɭ**, **r** vs. **ɽ**).

This area with languages exhibiting at least a twofold retroflex contrast includes some Indo-Aryan languages (Hindi, Oriya and Standard Bengali), the Dravidian languages,⁸ and the Munda languages, with two exceptions: (1) In

⁷ A precise definition is given by Laver (1994:141): “curling it [= the tongue tip] so that it points either upwards or slightly backwards, presenting either the tip or the undersurface of the tip as the active articulator to some part of the palate as the passive articulator.” Ladefoged and Maddieson (1996:27) distinguish between apical-postalveolar-retroflex, written as **ɽ**, occurring e.g., in Hindi, and sub-apical-palatal retroflex, written as **ɽ** and occurring, e.g., in Tamil and Telugu.

⁸ Telugu, which is, however, outside our area, is also a special case: it has no opposition **t** vs. **ɽ** (Dravidian **t** has become **θ**), but since the alveolar fricative, nasal and trill have a retroflex counterpart (**ʃ**, **ɳ**, **ɽ**), the feature “retroflex” is surely present. According to Pinnow

Korku the only retroflex phoneme is ɽ . (2) The superposed number for Sora (So¹) stands for total absence of distinctive retroflex sounds.⁹ This is the only language without any dental-retroflex opposition in the western part of the area under investigation.

North of Nepali and Maithili and including these, the feature retroflex is often confined to a single opposition, namely, in the Tibeto-Burman languages Tamang, Gurung and Thakali.

A special case is represented by some Tibetan dialects which have, like Chinese, a retroflex affricate tʂ (vs. ts) and a retroflex fricative ʂ (vs. s). This is a different type of retroflex pattern than common South Asian t vs. ɽ . It may be noted that other Tibetan dialects have the opposition t vs. ɽ (e.g., Lhasa Tibetan and Jirel), but a distinctive retroflex liquid that is typical for the southern area never occurs.

In the west the retroflex area extends to some of the Iranian languages, in both the Western and Eastern groups (e.g., Pashto, Baluchi, Pamir languages like Wakhi and Sanglechi, and the isolate Burushaski. Other Iranian languages (e.g., Persian, Ossetic), however, have no retroflexes. The neighbouring language family to the north, Altaic, has no retroflexes at all.

In the remaining languages to the east, i.e. mainly Tibeto-Burman (up to Kiranti in the west), Assamese, Chittagong Bengali and Southeast Asian language families, retroflex sounds generally play no distinctive role. Exceptions are Hmong and Vietnamese: Hmong is a rare case in Southeast Asia in having retroflexion involved in three oppositions (t vs. ɽ , ts vs. tʂ , s vs. ʂ). Southern Vietnamese dialects have retroflex phonemes as well, but they are in the process of disappearing.¹⁰

Other deviations from the typical case are marginal. There are languages where the only stops in the coronal domain are retroflex or postalveolar (e.g., Athpare). Other languages use retroflex sounds only in loan-words (e.g., Limbu, Chamling).

According to historical sources and investigations¹¹ it is generally accepted that the retroflex sounds in Indo-Aryan have been borrowed from Dravidian (see Kuiper 1967 for evidence). Since they are missing in the reconstructed proto-languages for Austro-Asiatic and Tibeto-Burman, I assume that a

(1964:48) the phonemic status of p and ɽ is limited, so that there may remain a single dental-retroflex opposition in the stops.

⁹ According to Stampe (1965:333) Sora has t and ɸ , but no ɽ or d .

¹⁰ I owe this information to Brigitt Gutmann.

¹¹ My sources for historical material are Zvelebil (1990:7) for Dravidian, Pinnow (1963:147f) and (1959:427) for Austro-Asiatic, DeLancey (1987:76) and Benedict (1972:17) for Tibeto-Burman, and Masica (1991:157f) and Burrow (1973:67f) for Indo-Aryan.

borrowing process has taken place here as well, either directly from Dravidian or via Indo-Aryan.

2.2. *Paucity of fricatives*

Another striking feature of the phoneme inventories in our area is the relatively small number of fricatives. We can exclude here the glottal “fricative” *h* since it does not correspond to the conventional definition of fricatives as being produced with turbulent airstream within the vocal tract (Ladefoged and Maddieson 1996:137).

Languages with only one fricative are comparatively rare in the world, as Maddieson’s (1984) calculation from a sample of 317 languages shows: 11.7%. This single fricative is most commonly /s/ (83% of all languages have an *s*-sound, as do 88.5% of the languages with fricatives, or 31 of 37 languages with only one fricative). The findings in our area conform to this universal tendency, with two exceptions: Parji, having only *v*, and Bengali, having only *ʃ*. Languages with two fricatives are more frequent: they are represented by a percentage of 19.6 in Maddieson’s study.

The isogloss of fricative paucity (that is two or less fricatives) delimits a sector that is almost identical to the area in question. Excluded are some of the Tibeto-Burman languages: some Tibetan dialects¹², Lepcha, Dzongkha, Mizo, and Mao Naga. The remaining languages have one or two fricatives: the Indo-Aryan languages (except Assamese and Chittagong Bengali), the Munda languages (all having only one), Khasi, the Dravidian languages¹³, and most of the Tibeto-Burman languages. About two thirds of those have only one fricative.

Languages in the eastern adjacent area show considerably more fricatives (e.g., Lisu has six; a counterexample is Mon with only one). This feature is shared by some of the most eastern languages in our area, Mizo and Mao Naga (but not by Manipuri). In the west small sets of fricatives are still found, extending to Marwari at the border of Pakistan. Persian, however, has an elaborated system of fricatives. In the south Telugu marks a different type, containing five fricatives.

Since all proto-languages are assumed to have few fricatives (Tibeto-Burman *s* and *z*, Indo-Aryan *s*, *ʃ*, *ʒ*¹⁴) or no fricatives at all (Dravidian and

¹² Lhasa Tibetan has only two fricatives, *s* and *ʃ* (Goldstein and Nornang 1978:xiv), but many of the other dialects have more (Namkung 1996:387f).

¹³ As described by Mahapatra (1979:207), Malto has three major dialects, of which one (the Kumarbhag dialect) has only one fricative, the other two (the Malpaharia dialect and the Sawriya dialect) have three fricatives.

¹⁴ OIA has three fricatives. They are, however, very likely to coalesce for reasons of phonetic similarity.

Austro-Asiatic), it is not possible to give any direction of spreading for that property.

2.3 *Breathy voice*

The next two traits are both related to a glottal feature. Breathy voice on the one hand and creaky voice on the other may be integrated in a continuum (Ladefoged and Maddieson 1996:49) that ranges from wide opening to complete closure of the glottis. Breathy voice shows an opening that is wider than that of voiced sounds (modal voice in Ladefoged and Maddieson's terms) but narrower than that of voiceless sounds, whereas the vocal cords in creaky voiced sounds are much more closed than in modal voice, but not pressed together as for the glottal stop.¹⁵

The term "breathy voice" hardly occurs in any of the descriptions of the languages in the area. Most often the phenomenon is assigned to aspiration since the breathy voiced sounds are in the same relationship to the voiced plain ones as are the aspirated voiceless sounds to the voiceless plain ones (e.g., **b^h** vs. **b** / **p^h** vs. **p**). Both "aspirated" series share some phonetic features, such as the relatively high rate of oral airflow or the rapid and brief drop in subglottal air pressure (see Ladefoged and Maddieson 1996:58).

Breathy voice is rather rare in the world's languages. Besides South Asia it is found in some African languages such as Igbo, Shona, Zulu, Ndebele and Zhul'oasi (Khoisan).

The isogloss of the feature "breathy voice" cuts across the area of interest, excluding Assam and most of the eastern part of our area (with two exceptions), but with an extension in the west. Within this region, an opposition between voiced vs. breathy voiced plosive is found (e.g., **b** vs. **b^h**). Included are all Indo-Aryan languages except the most western ones (Kashmiri and Punjabi); among the Munda languages the northern group¹⁶ and Kharia; Gondi and both dialects of Kurux as Dravidian representatives; and some of the more western Tibeto-Burman languages (Gurung, Tamang¹⁷, Newari, several Kiranti languages and Dzongkha).

Exceptions are Bodo and Garo, where the aspirates have only allophonic status; Khasi, which (unlike most Austro-Asiatic languages) has voiceless

¹⁵ For my use of the term "creaky voice", see 2.4.

¹⁶ Cook (1965:4) reports that the two principal dialects of Mundari differ from each other in the status of aspiration. The Naguri dialect has phonemic aspiration, while in the Hasada dialect the aspiration is non-phonemic. Ghosh (1992:17) notes for Santali that aspirates occur only in loanwords, in contrast to Pinnow (1959:36), who does not mention such a restriction.

¹⁷ For Tamang I have found contradictory accounts: Mazaudon (1973) does not mention any voiced stop, whereas Weidert (1987:261) reports breathy voice. On the two maps Weidert's account is shown.

aspirated plosives; Malto, which has **h** as a phoneme, but otherwise no aspirated sounds, and Juang and Sora, which have no aspiration at all (neither has a phoneme **h**). In the south, the Dravidian languages from Telugu and Parji southwards are excluded.

Up to now we have been concerned with plosives. In many of the languages mentioned above breathy voice is also found with nasals and/or liquids. On Map 1 languages with breathy voiced nasals are marked by underlined **f̥**: this applies to Bengali¹⁸, Bhojpuri, Nepali, Newari, Chamling, Dhimal and Dhangar Kurux.

A related phenomenon is found in Angami (a Naga language located in our area, but not included in our sample), which has voiceless nasals. They may be described as being aspirated because the timing relationship of the glottal opening after the oral closure is the same as in aspirated stops (Ladefoged and Maddieson 1996:116).

Other voiceless nasals and liquids reported in some Southeast Asian languages (e.g., Burmese, and probably Hmong, Mon, and Palaung) and in the Hmar dialect of Mizo (Ladefoged and Maddieson 1996:113) are clearly different. There, some voicing exists just before the articulators come apart, i.e. the vocal folds tighten instead of widen as in aspiration (p.111). They are often transcribed by “**h+C**” which may be easily confused with the transcription for breathy voiced segments “**C+h**”. Functionally (as in causativization by aspiration of the initial consonant), however, they are parallel to the voiceless aspirated plosives (Ladefoged and Maddieson 1996:69). On Map 1 they are designated by underlined **h**.

It might be supposed that aspiration of voiceless segments is much more frequent. This is not the case, as is shown by the percentage given in Ruhlen (1976:155): only 13%¹⁹ of the languages in his sample attest an aspiration contrast in stops. Thus it may serve as an areal feature, not for the area in question, but for South and Southeast Asia, since there is no such concentration of languages with aspiration elsewhere in the world. Exceptions inside the area are few: Chittagong Bengali, Bodo and Garo, Apatani, Malto, Juang and Sora.

The only language family for which breathy voiced consonants are postulated in the ancestor language is Indo-Aryan. Proto-Indo-European is reconstructed as having a voiced series of aspirated plosives. Pinnow (1959)

¹⁸ See Ferguson and Chowdhury (1960:46). They note aspirated sonorants (**m**, **n**, **l**, **r**, **ɽ**) in the speech of educated speakers of Standard Colloquial Bengali. On the other hand, the dialect of Chittagong lacks aspiration altogether (Ucida 1970:8).

¹⁹ This number must be too low as his sample is biased towards America and Europe, Asian languages being under-represented. From the percentages given in Maddieson (1984:26f) I calculated a rate of about 20% of his sample having an aspiration contrast in stops.

claims a development of aspiration from early Proto-Munda to late Proto-Munda. The most plausible source for this feature is Indo-Aryan.

2.4. *Creaky voice and unreleased final consonants*

The investigation of creaky voice is complicated by two factors. Firstly, it is interrelated with other phenomena that are not easily distinguished from it: checked vowels, (pre-)glottalised consonants, and unreleased consonants. In creaky voiced sounds the vocal cords are much closer than in voiced ones.²⁰ The glottal constriction may occur before the oral closure, resulting in what may be called a preglottalised sound. Checked vowels are reported for languages in Southeast Asia. They are defined by a following glottal closure. The feature “unreleased” is added here since preglottalisation may prevent the release of the oral closure, as I noticed in my own research on Juang. I suppose that the unreleased consonants of some languages in the area are at the same time glottalised, a fact that may go unnoticed because it is difficult to hear. Unreleased closure and glottalisation, however, are not necessarily connected, as is shown in English, where we have unreleased non-glottalised stops like *k* in [stakt̚] ‘stacked’ (Laver 1994:359). For the purpose of this study the sounds described by these labels are all subsumed under the feature “creaky voice”.

Secondly, I propose to examine not only the phonemic but also the allophonic use of glottalised consonants. Reference to such allophony is often even more difficult to find in the literature, if it is not absent altogether. I restrict my study to occurrences of glottalised consonants in word-final position.

The feature “creaky voice” is marked by outlined script on Map 2. It is found throughout the area in question and in neighbouring languages, but by no means throughout the whole area.

Allophonic use of glottalisation occurs more commonly in the western part of the area. As described by Pinnow (1959), Munda languages have (pre)glottalised unreleased final consonants. Santali has non-glottalised final stops as well, but they appear to occur in Hindi loan-words only (Zide 1958:46). The situation in Kharia (Biligiri 1965:14) and in Mundari (Osada 1992:28) is unclear: phonemic contrasts can be found occasionally between glottalised and plain voiced plosives. According to Rabel (1961:2) final stops in Khasi are unreleased, whereas the closely related Amwi (also a Khasian

²⁰ This description follows Ladefoged and Maddieson (1996:53f). The alternative term “laryngealised consonants” is also covered by their definition of creaky voice. I exclude here the implosives, which are widely attested in Southeast Asia (for details and a historical perspective see Solnit 1992) and sporadically in Indo-Aryan languages (Marwari, Sindhi). The implosive consonants show a different distribution: they occur mainly in syllable-initial position.

language) is said to have final plosives that are also glottalised (Weidert 1975b:11f). Dhangar Kurux in Nepal is reported to have final consonants with delayed release (Gordon 1973:39), but Pinnow (1964) in his detailed article about the phonology of the somewhat different Kurux dialect in South Bihar does not mention any such phenomenon. At least the glottal stop (which is phonemic in no other Dravidian language) is sure to be present (e.g., many verbs end in -ʔ). The Tibeto-Burman languages Thulung, Athpare, Limbu, Dhimal and Mikir and the Tai language Khamti have unreleased final consonants, at least in Limbu with a simultaneous glottal closure, as mentioned explicitly by van Driem (1987:2). Unexploded final consonants occur in Tai languages (Henderson 1965:428), and according to Diffloth and Zide (1992:140) "this is the phonetic norm across language families in the Southeast Asian area, but it stands out as a characteristic feature of the Munda languages in the Indian linguistic area."

Phonemic use of creaky voice is widespread in Southeast Asia and in Chinese dialects (Sagart 1989): Burmese, Thai, Karen have it, to cite just a few.²¹ Only some scattered occurrences are found in our area. Bodo, Garo and Mizo (all Tibeto-Burman) have a series of glottalised final consonants (glottalised nasals in Bodo and Garo and glottalised liquids in Mizo). Other oral final consonants are not released (Bodo, Garo) or are absent altogether (Mizo). In addition, a phonemic²² glottal stop in syllable-final position (in opposition to syllables ending in a vowel) occurs in these three languages, as well as in Apatani, Nocte, Jingpho, and Mikir.

It is obvious from the discussion so far that the creaky voice phenomenon is absent in the Indo-Aryan and Dravidian languages. One exception in the area is Chittagong Bengali, which is reported to have unreleased variants for voiceless plosives in final position (Ucida 1970:8).

It is not easy to make statements about the phonemic occurrence of these segments in the world's languages. In Maddieson's (1984:205f.) sample, 9 of 371 languages contain a glottalised (in his terms laryngealised) plosive: two African languages (Logbara, Ngizim), three Amerindian languages (Otomi, K'ekchi, Wapishana) and four Southeast Asian languages (Sedang, Lakkia, Sui, Lungchow).²³ Two languages are reported to have contrastively laryngealised vowels, Nambiquara (Amerindian) and Sedang (Austro-Asiatic). As long as no other occurrences are reported, it can be assumed that the

²¹ It should be noted that some of these languages may be better described as having stiff voice (i.e. sounds of an intermediate type between modal voice and creaky voice) instead of creaky voice, as explicitly stated for Thai by Ladefoged and Maddieson (1996:55).

²² But often interdependent with tonal distinctions.

²³ The distribution of the related implosive consonants, attested in 41 of 317 languages, shows a clearer picture with a strong concentration in Africa (29 of 41 languages).

concentration in Southeast Asia is unique. Here we have the feature “final glottalised consonants”, characterizing a large region; our area seems to lie on its western fringes.

The distribution of phonetic glottal phenomena found in the area is even more difficult to interpret since typological studies are missing altogether, as far as I am aware. It could be a common feature of the languages in the western part of our area.

From the historical point of view two hypotheses may be reported. Sagart (1989) assumes checked syllables for Proto-Sino-Tibetan, and Pinnow (1959:427) claims implosive or non-released consonants for late Proto-Munda. Hence we might very tentatively suggest a spreading process from Tibeto-Burman in the east by various intermediate stages to Munda and some adjacent languages in the west.

2.5. *Initial velar nasal*

Phonemic velar nasals are far from rare, but they are less frequent than bilabial or dental/alveolar nasals (Maddieson 1984:60). If a language has three nasals, then one of them is velar (166 of 199 languages in his sample). The distribution of the opposition of alveolar vs. velar nasals is not shown on Map 2, but a look at the cases in Map 1 reveals that only very few languages (Newari and the Indo-Aryan languages Hindi, Oriya and Maithili) lack this contrast.²⁴ The remainder of the languages exhibit distinctive η . Outside our area phonemic η is missing in the adjacent languages in the west and in the south (Telugu, Kannada, Marathi, Marwari).

Thus an isogloss of phonemic velar nasals would neatly delimit the area in question, if there were not two facts to consider: (1) the distribution of η continues into Southeast Asia and (2) the relative frequency of velar nasals in general.

As the presence or absence of a phonemic velar nasal in a group of languages is therefore not very significant, we shall be concerned mainly with its distribution in the syllable pattern. It is interesting that the impressionistically frequent constraint to avoid initial velar nasals is not valid in all the languages of our area. Unfortunately, there is no typological survey on the distribution of nasals. I know of a few incidences of initial velar nasals in languages scattered over the world: they occur in Bantu (e.g., Swahili, Venda) and Chadic (Kera), marginally in Celtic languages (e.g., Irish, Welsh),²⁵ and widely in Australian languages (Dixon 1980).

²⁴ The velar nasal is always present as an allophonic variant before velar plosives.

²⁵ Initial velar nasals, however, do not occur in sentence-initial position.

The solid line on Map 2 shows the presence of an initial velar nasal, absent from large parts of the area of interest. Initial velar nasals are then much rarer, though they are widespread in Southeast Asia. In our area Khasi and almost all of the Tibeto-Burman languages show them, exceptions being Bodo, Garo, Mikir, Dhimal and Newari.

After the glottalised consonants this is the second feature that has its centre in Southeast Asia and spreads towards the west, with an even stronger representation.

An initial velar nasal was most likely present in Proto-Tibeto-Burman and in Proto-Austro-Asiatic, whose descendants are the only languages where initial η is attested. Instead of hypothesising the emergence of initial η in some languages, we can assume a converse development: the Munda languages have lost this initial occurrence,²⁶ maybe through Indo-Aryan influence.

3. SYNTHESIS

Table 2 lists the features with their distribution inside and outside the area in question.²⁷ Included is the postulated direction of the spreading and its possible origins.

Looking at Map 2, it is clear at first sight that, taken together, the isoglosses for all five features do not combine to make a single bundle. Neither the area suggested by Ebert (1993) nor the extended area under study in the current paper is clearly delimited. Nevertheless a sort of pattern is visible.

Three features occur mainly in languages in the west, two primarily in those in the east. In both cases the features extend beyond the borders of our area: while the western group shares retroflexes and breathy voice with many languages in the western and central parts of India, languages of Southeast Asia (including, e.g., Burmese) have creaky voice and initial velar nasals in common with the eastern group. The paucity of fricatives is the only feature that is characteristic for the whole area, but also for adjacent Central India.

In the case of the retroflexes and the breathy voiced sounds, historical evidence supports the hypothesis of an eastward spreading of these two features. As for the presence or absence of initial velar nasals it is more plausible to assume a converse process in the eastward direction: initial velar nasals seem to have disappeared under the influence of Indo-Aryan or Dravidian. The only influence in the westward direction is possibly the spreading of creaky voice.

²⁶ E.g., Mundari replaced it by n : *nur* 'to flow out' < ** η ur* (Pinnow 1959:211).

²⁷ Languages mentioned by name constitute the boundaries where the feature no longer occurs.

In sum, our area constitutes a transition zone from (Central) India to Southeast Asia, with more languages showing phonetic features of the South Asian type than of the Southeast Asian type.

<i>Features</i>	<i>distribution inside the area</i>	<i>distribution outside the area</i>	<i>postulated spreading</i>	<i>possible sources</i>
1. retroflexes	Western	South Asia	eastwards	IA
2. few fricatives	Western, central	South Asia (N: Tibetan, W: Pashto)	no spreading	all
3. breathy voice	Western	Central India (S: Telugu, W: Punjabi)	eastwards	IA
4. creaky voice	Eastern	Southeast Asia	westwards	TB
5. absence of initial velar nasal	Western	Southeast Asia	eastwards	IA, DR

Table 2. Summary

APPENDIX I
Alphabetic list of languages considered

(AA = Austro-Asiatic, DR = Dravidian, IE = Indo-European, MI = Miao-Yao, SI = Sinitic, TA = Tai, TB = Tibeto-Burman)

1.	Apatani	Ap	TB	(Weidert 1987)
2.	Assamese	As	IE	(Kakati 1941)
3.	Athpare	At	TB	(Ebert 1997b)
4.	Bengali, Standard	Be	IE	(Ferguson & Chowdhuri 1960)
5.	Bhojpuri	Bh	IE	(Shukla 1981)
6.	Bodo	Bo	TB	(Burling 1959)
7.	Burmese	Bu	TB	(Okell 1969; Matthias Jenny, p.c.)
8.	Chamling	Ca	TB	(Ebert 1997a)
9.	Cantonese		SI	(Matthews & Yip 1994)
10.	Chinese, Mandarin		SI	(Li & Thompson 1981)
11.	Chittagong Bengali	Ch	IE	(Ucida 1970)
12.	Chrau		AA	(Thomas 1971)
13.	Dai		TA	(Hartmann 1984)
14.	Dhangar Kurux	Dk	DR	(Gordon 1973)
15.	Dhimal	Dh	TB	(K. Cooper, p.c.; King 1994)
16.	Dzongkha	Dz	TB	(Mazaudon & Michailovsky 1989)
17.	Garo	Ga	TB	(Burling 1961)
18.	Gondi	Go	DR	(Subrahmanyam 1968)
19.	Gurung	Gu	TB	(Glover 1974)
20.	Hindi	Hi	IE	(Fairbanks & Misra 1966; Masica 1991)
21.	Hmong		MI	(Harriehausen 1990; Mottin 1978)
22.	Jingpho	Ji	TB	(Namkung 1996)
23.	Juang	Ju	AA	(own research; Pinnow 1959)
24.	Kannada		DR	(Schiffman 1983)
25.	Kashmiri		IE	(Wali & Koul 1997)
26.	Khamti	Km	TA	(Weidert 1977)
27.	Kharia	Kr	AA	(Biligiri 1965)
28.	Khasi	Ks	AA	(Rabel 1961)
29.	Khmer		AA	(Jacob 1968)
30.	Korku	Ko	AA	(Zide 1960)

31. Kurux	Ku	DR	(Pinnow 1964; Pfeiffer 1972)
32. Lahu	La	TB	(Matisoff 1988)
33. Lepcha	Le	TB	(Mainwaring 1876; Sprigg 1989)
34. Limbu	Li	TB	(van Driem 1987)
35. Lisu, Northern	Ls	TB	(Bradley 1994)
36. Maithili	Mt	IE	(Yadav 1996)
37. Malto	Ml	DR	(Mahapatra 1979)
38. Manipuri	Mn	TB	(Chelliah 1997)
39. Mao-Naga	Ma	TB	(Giridhar 1994)
40. Marathi	Mr	IE	(Pandharipande 1997)
41. Marwari (Rajasthani)		IE	(Masica 1991)
42. Mikir	Mk	TB	(Grüssner 1978)
43. Mizo	Mz	TB	(Weidert 1975a)
44. Mon	Mo	AA	(Matthias Jenny, p.c.; Pinnow 1959)
45. Mundari	Mu	AA	(Cook 1965; Osada 1992)
46. Nepali	Ne	IE	(Masica 1991)
47. Newari	Nw	TB	(Malla 1985)
48. Nocte	No	TB	(Das Gupta 1971; French 1983)
49. Oriya	Or	IE	(own research; Masica 1991)
50. Palaung	Pa	AA	(Shorto 1960)
51. Parji	Pj	DR	(Burrow & Bhattacharya 1953)
52. Pashto		IE	(Shafeev 1964)
53. Persian		IE	(Lambton 1976)
54. Punjabi		IE	(Bhatia 1993)
55. Santali	Sa	AA	(Pinnow 1959; Ghosh 1992)
56. Sgaw / Karen	Sg	TB	(Jones 1961)
57. Sindhi		IE	(Masica 1991)
58. Sora	So	AA	(Stampe 1965)
59. Tamang	Ta	TB	(Mazaudon 1973; Weidert 1987)
60. Telugu	Te	DR	(Lisker 1963)
61. Thai		TA	(Matthias Jenny, p.c.; Noss 1964)
62. Thulung	Th	TB	(Allen 1975)
63. Tibetan (Lhasa)	Ti	TB	(Goldstein & Nornang 1978)
64. Vietnamese		AA	(Nguy�en �inh-ho�a 1997)

APPENDIX II
Genetic classification of the languages surveyed

(* = part of the area under investigation)

Austro-Asiatic
Munda

North: Korku, Mundari*, Santali*
 Central: Juang*, Kharia*
 South: Sora*

Mon-Khmer

North: Khasi*, Palaung
 East: Khmer, Chrau
 Mon
 Vietnamese

Tai: Dai, Khamti*, Thai

Miao-Yao: Hmong

Sino-Tibetan

Sinitic: Cantonese, Mandarin

Tibeto-Burman

Bodish: Dzongkha*, Gurung*, Lhasa Tibetan*, Newari*,
 Tamang*

Eastern Himalayan: Athpare*, Camling*, Dhimal*, Limbu*,
 Thulung*

Bodo-Garo: Bodo*, Garo*

Konyak: Nocte*

Naga: Mao-Naga*

Kuki-Chin: Lepcha*, Mizo*

Mikir-Meithei: Manipuri*, Mikir*

Abor-Miri-Dafla: Apatani*

Kachinic: Jingpho

Lolo-Burmese: Bama (Burmese), Lahu, Lisu (Northern)

Karenic: Sgaw

Indo-European

Indo-Aryan (geographical classification)

Northern: Nepali*

Central: Bhojpuri*, Hindi, Maithili*, Marathi

Eastern: Assamese*, Bengali (Standard colloquial)*, Bengali
 (Chittagong)*, Oriya*

Western: Kashmiri, Marwari (Rajasthani), Punjabi, Sindhi

Iranian: Farsi, Pashto

Dravidian

North: Kurux*, Dhangar Kurux*, Malto*

Central: Gondi*, Parji*, Telugu

South: Kannada

BIBLIOGRAPHY

- ALLEN, Nicholas J. 1975. *Sketch of Thulung Grammar*. Ithaca: Cornell University.
- BENEDICT, Paul K. 1972. *Sino-Tibetan. a Conspectus*. Cambridge: Cambridge University Press.
- BHAT, D. N. S. 1973. "Retroflexion: an areal feature." *Working Papers on Language Universals* 13:27-67.
- BHATIA, Tej K. 1993. *Punjabi*. London: Routledge.
- BILIGIRI, Hemnige Shrinivasarangachar. 1965. *Kharia. Phonology, Grammar and Vocabulary*. Building Centenary and Silver Jubilee Series 3. Poona: Deccan College.
- BRADLEY, David. 1994. *A Dictionary of the Northern Dialect of Lisu (China and Southeast Asia)*. Pacific Linguistics Series C 126. Canberra: Australian National University.
- BURLING, R. 1959. "Proto-Bodo." *Language* 35.3:433-53.
- _____. 1961. *A Garo Grammar*. Deccan College Monograph Series 25. Poona: Deccan College, Postgraduate and Research Institute.
- BURROW, T. 1973. *The Sanskrit language*. London: Faber and Faber.
- _____, & S. BHATTACHARYA. 1953. *The Parji Language*. Hertford: Stephen Austin & Sons.
- CHELLIAH, Shobhana L. 1997. *A Grammar of Meithei*. Berlin: Mouton de Gruyter.
- COOK, Walter A. 1965. *A Descriptive Analysis of Mundari*. Ph.D. dissertation, Georgetown University (Washington, D.C.).
- DAS GUPTA, K. 1971. *An Introduction to the Nocte Language*. Shillong: North-East Frontier Agency.

- DeLANCEY, Scott. 1987. "Sino-Tibetan languages." *The Major Languages of East and South-East Asia*, ed. by B. Comrie, 71-82. London: Routledge.
- DIFFLOTH, Gérard, & Norman H. ZIDE. 1992. "Austro-Asiatic languages." *International Encyclopedia of Linguistics*, ed. by W. Bright, Vol. 1, 137-42. New York, Oxford: Oxford University Press.
- DIXON, R. M. W. 1980. *The Languages of Australia*. Cambridge: Cambridge University Press.
- DRIEM, George van. 1987. *A Grammar of Limbu*. Berlin: Mouton de Gruyter.
- EBERT, Karen H. 1993. "Kiranti subordination in the South Asian areal context." *Studies in Clause Linkage*, (Arbeiten des Seminars für Allgemeine Sprachwissenschaft 13.) ed. by K. H. Ebert, 83-110. Zürich: Universität.
- _____. 1997a. *Camling (Chamling)*. Languages of the World/ Materials 103. München: Lincom.
- _____. 1997b. *A Grammar of Athpare*. Lincom Studies in Asian Linguistics 1. München: Lincom.
- FAIRBANKS, Gordon H., & Bal Govind MISRA. 1966. *Spoken and Written Hindi*. Ithaca: Cornell University Press.
- FERGUSON, Charles A., & Munier CHOWDHURY. 1960. "The phonemes of Bengali." *Language* 36.1:22-59.
- FRENCH, Walter Thomas. 1983. *Northern Naga: a Tibeto-Burman Mesolanguage*. 2 volumes. Ph.D. dissertation, City University of New York.
- GHOSH, Arun. 1994. *Santali: a Look into Santal Morphology*. New Delhi: Gyan Publishing House.
- GIRIDHAR, P. P. 1994. *Mao Naga Grammar*. Mysore: Central Institute of Indian Languages.

- GLOVER, Warren W. 1974. *Sememic and Grammatical Structures in Gurung (Nepal)*. Norman: Summer Institute of Linguistics.
- GOLDSTEIN, Melvyn C., & Nawang NORNANG. 1978. *Modern Spoken Tibetan: Lhasa Dialect*. Bibliotheca Himalayica, Series II, 14. Kathmandu: Ratna Pustak Bhandar.
- GORDON, Kent H. 1973. "Clause patterns in Dhangar-Kurux." *Patterns in Clauses, Sentence, and Discourse in Selected Languages of India and Nepal*, ed. by R. L. Trail, Vol. II: Clause, 37-122. Kathmandu: Tribhuvan University Press.
- GRIERSON, George A. 1903-28 [1968]. *Linguistic Survey of India*. Calcutta (Repr. Delhi: Motilal Banarsidass).
- GRÜSSNER, Karl-Heinz. 1978. *Arleng Alam, die Sprache der Mikir*. Beiträge zur Südasiensforschung 39. Wiesbaden: Steiner.
- HARRIEHAUSEN, Bettina. 1990. *Hmong Njua*. Tübingen: Niemeyer.
- HARTMANN, John F. 1984. *Linguistic and Memory Structures in Tai-Lue Oral Narratives*. Pacific Linguistics, Series B 90. Canberra: Australian National University.
- HENDERSON, Eugenie J. A. 1965. "The topography of certain phonetic and morphological characteristics of South East Asian languages." *Lingua* 15:400-34.
- JACOB, Judith M. 1968. *Introduction to Cambodian*. London: Oxford University Press.
- JONES, Robert B. 1961. *Karen Linguistic Studies*. Berkeley: University of California.
- KAKATI, Banikanta. 1962. *Assamese, Its Formation and Development*. Gauhati: Lawyer's Book Stall.
- KING, John. 1994. "Picking up where Hodgson left off: further notes on Dhimal." *Linguistics of the Tibeto-Burman Area* 17.2:121-32.

- KUIPER, F. B. J. 1967. "The genesis of a linguistic area." *Indo-Iranian Journal* 10:81-102.
- LADEFOGED, Peter, & Ian MADDIESON. 1996. *The Sounds of the World's Languages*. London: Blackwell.
- LAMBTON, Ann K. S. 1976. *Persian Grammar*. Cambridge: Cambridge University Press.
- LAVER, John. 1994. *Principles of Phonetics*. Cambridge: Cambridge University Press.
- LI, Charles N., & Sandra A. THOMPSON. 1981. *Mandarin Chinese. a Functional Reference Grammar*. Berkeley: University of California Press.
- LISKER, Leigh. 1963. *Introduction to Spoken Telugu*. New York: American Council of Learned Societies.
- MADDIESON, Ian. 1984. *Patterns of Sounds*. Cambridge: Cambridge University Press.
- MAHAPATRA, B. P. 1979. *Malto: an Ethno-Semantic Study*. Mysore: Central Institute of Indian Languages.
- MAINWARING, G. B. 1876 [1971]. *A Grammar of the Lepcha Language*. Bibliotheca Himalayica, Series II 5. New Delhi: Manjusri Publishing House.
- MALLA, Kamal P. 1985. *The Newari Language: a Working Outline*. Monumenta Serindica 14. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.
- MASICA, Colin P. 1976. *Defining a Linguistic Area: South Asia*. Chicago: University of Chicago Press.
- _____. 1991. *The Indo-Aryan Languages*. Cambridge: Cambridge University Press.

- MATISOFF, James A. 1988. *The Dictionary of Lahu*. University of California Publications in Linguistics 111. Berkeley: University of California Press.
- MATTHEWS, Stephen, & Virginia YIP. 1994. *Cantonese*. London: Routledge.
- MAZAUDON, Martine. 1973. *Phonologie tamang*. Paris: Société des Études Linguistiques et Anthropologiques de France (SELAF).
- _____, & Boyd MICHAILOVSKY. 1989. "Lost syllables and tonal contour in Dzongkha." *Prosodic Analysis and Asian Linguistics: to Honour R. K. Sprigg*, ed. by D. Bradley, E. J. A. Henderson, & M. Mazaudon, 115-36. Canberra: Australian National University.
- MICHAILOVSKY, Boyd. 1988. "Phonological typology of Nepal languages." *Linguistics of the Tibeto-Burman Area* 11.2:25-50.
- MOTTIN, Jean. 1978. *Éléments de grammaire hmong blanc*. Bangkok: Don Bosco Press.
- NAGARAJA, K. S. 1985. *Khasi*. Pune: Deccan College, Post-Graduate and Research Institute.
- NAMKUNG, Ju (ed.) 1996. *Phonological Inventories of Tibeto-Burman Languages*. STEDT Monograph Series 3. University of California, Berkeley: Center for Southeast Asia Studies.
- NGUYEN Ính-hoà. 1997. *Vietnamese*. London Oriental and African Language Library 9. Amsterdam: Benjamins.
- NOSS, Richard B. 1964. *Thai Reference Grammar*. Washington: Foreign Service Institute.
- OKELL, John. 1969. *A Reference Grammar of Colloquial Burmese*. 2 vols. London: Oxford University Press.
- OSADA, Toshiki. 1992. *A Reference Grammar of Mundari*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.

- PANDHARIPANDE, Rajeshwari V. 1997. *Marathi*. London: Routledge.
- PFEIFFER, Martin. 1972. *Elements of Kurux Historical Phonology*. Leiden: E. J. Brill.
- PINNOW, Heinz-Jürgen. 1959. *Versuch einer historischen Lautlehre der Kharia-Sprache*. Wiesbaden: Harrassowitz.
- _____. 1963. "The position of the Munda languages within the AustroAsiatic language family." *Linguistic Comparison in South East Asia and the Pacific*, ed. by H. L. Shorto, 140-52. School of Oriental and African Studies, University of London.
- _____. 1964. "Bemerkungen zur Phonetik und Phonemik des Kurukh." *Indo-Iranian Journal* 8.1:32-59.
- RABEL, Lili. 1961. *Khasi, a Language of Assam*. Baton Rouge: Louisiana State University Press.
- RAMANUJAN, A. K., & Colin P. MASICA. 1969. "Toward a phonological typology of the Indian linguistic area." *Current Trends in Linguistics*, ed. by T. A. Sebeok, Vol. 5: Linguistics in South Asia, 543-77. The Hague: Mouton.
- RUHLEN, Merritt. 1976. "The geographical and genetic distribution of linguistic features." *Linguistic Studies Offered to Joseph Greenberg*, ed. by A. Juillard, Vol. 1, 137-60. Saratoga: Anma Libri.
- SAGART, Laurent. 1989. "Glottalised tones in China and South-East Asia." *Prosodic Analysis and Asian Linguistics: to Honour R. K. Sprigg*, ed. by D. Bradley, E. J. A. Henderson, & M. Mazaudon, 83-93. Canberra: Australian National University.
- SCHIFFMAN, Harold F. 1983. *A Reference Grammar of Spoken Kannada*. Seattle: University of Washington Press.
- SHAFEEV, D. A. 1964. *A Short Grammatical Outline of Pashto*. Bloomington: Indiana University.

- SHORTO, H. L. 1960. "Word and syllable patterns in Palaung." *Bulletin of the School of Oriental and African Studies* 23:544-57.
- SHUKLA, Shaligram. 1981. *Bhojpuri Grammar*. Washington: Georgetown University Press.
- SOLNIT, David B. 1992. "Glottalized consonants as a genetic feature in Southeast Asia." *Acta Linguistica Hafniensa* 25:95-123.
- SPRIGG, R. K. 1989. "Oral vowels and nasalized vowels in Lepcha (Rong) as the key to a puzzling variation in spelling." *South-East Asian Linguistics: Essays in Honour of Eugénie J. A. Henderson*, ed. by J. H. C. S. Davidson, 219-35. School of Oriental and African Studies, University of London.
- STAMPE, David L. 1965. "Recent work in Munda linguistics I." *International Journal of American Linguistics* 31.4:332-41.
- SUBRAHMANYAM, P. S. 1968. *A Descriptive Grammar of Gondi*. Annamalainagar: Annamalai University.
- THOMAS, David D. 1971. *Chrau Grammar*. Oceanic Linguistics Special Publication 7. Honolulu: University of Hawaii Press.
- UCIDA, Norihiko. 1970. *Der Bengali-Dialekt von Chittagong*. Wiesbaden: Harrassowitz.
- WALI, Kashi, & Omkar N. KOUL. 1997. *Kashmiri*. London: Routledge.
- WEIDERT, Alfons. 1975a. *Componential Analysis of Lushai Phonology*. Amsterdam: Benjamins.
- _____. 1975b. *I tkong Amwi. Deskriptive Analyse eines Wardialekts des Khasi*. Wiesbaden: Harrassowitz.
- _____. 1977. *Tai-Khamti Phonology and Vocabulary*. Wiesbaden: Franz Steiner Verlag.
- _____. 1987. *Tibeto-Burman Tonology*. Current Issues in Linguistic Theory 54. Amsterdam: Benjamins.

YADAV, Ramawatar. 1996. *A Reference Grammar of Maithili*. Berlin: Mouton de Gruyter.

ZIDE, Norman H. 1958. "Final stops in Korku and Santali." *Indian Linguistics* 19 (Turner Jubilee volume 1):44-48.

_____. 1960. *Korku Phonology and Morphophonemics*. Ph.D. dissertation, University of Pennsylvania.

ZVELEBIL, Kamil V. 1990. *Dravidian Linguistics: an Introduction*. Pondicherry: Pondicherry Institute of Linguistics and Culture.