Prepublication draft. To appear as: Arsenault, Paul. 2015. Retroflex consonant harmony: An areal feature in South Asia. Journal of South Asian Languages and Linguistics 2(1). 1–28. http://dx.doi.org/10.1515/jsall-2015-0001.

Submission for Journal of South Asian Languages and Linguistics (JSALL)

Author: Paul Arsenault
Assistant Professor of Linguistics
Tyndale University College
25 Ballyconnor Court
Toronto, ON, Canada, M2M 4B3
parsenault@tyndale.ca

Short Title: Retroflex consonant harmony

Size: 51,402 characters (not counting spaces); 9,360 words

Abstract: Retroflexion is a well-known areal feature of South Asia. Most South Asian languages, regardless of their genetic affiliation, contrast retroflex consonants with their non-retroflex dental counterparts. However, South Asian languages differ in the phonotactic restrictions that they place on retroflex consonants. This paper presents evidence that a large number of South Asian languages have developed a co-occurrence restriction on coronal obstruents that can be described as retroflex consonant harmony. In these languages, roots containing two non-adjacent coronal stops are primarily limited to those with two dentals (T…T) or two retroflexes (Ṭ…Ṭ), while those containing a combination of dental and retroflex stops are avoided (*T…Ṭ, *Ṭ…T). Historical-comparative evidence indicates that long-distance retroflex assimilation has contributed to the development of this phonotactic pattern (T…Ṭ → Ṭ…Ṭ). In addition, the paper demonstrates that the distribution of languages with and without retroflex consonant harmony is geographic in nature, not genetic. Retroflex consonant harmony is characteristic of most languages in the northern half of the South Asian subcontinent, regardless of whether they are Indo-Aryan, Dravidian or Munda (but not Tibeto-Burman). It is not characteristic of Indo-Aryan and Dravidian languages in the south. Thus, retroflex consonant harmony constitutes an areal feature within South Asia.

Keywords: retroflex; consonant harmony; phonotactics; assimilation; areal feature

Retroflex consonant harmony: An areal feature in South Asia

1. Introduction

South Asia has long been recognized as a *linguistic area*, a geographic region in which languages of different genetic stock have come to resemble one another through a long history of contact and convergence. From a phonological point of view, the most prominent areal trait of South Asia is retroflexion. Most South Asian languages, regardless of their genetic affiliation, contrast retroflex consonants with non-retroflex dental or alveolar counterparts (Emeneau 1956; Ramanujan & Masica 1969; Bhat 1973). Retroflexion is reported in all Dravidian languages, all Indo-Aryan languages but Asamiya, all Munda languages but Sora, and a minority of Tibeto-Burman languages. Every South Asian language that maintains an opposition between retroflex and non-retroflex consonants maintains it between stops. Many languages extend the opposition to nasals and liquids, and a few extend it to fricatives, affricates or approximants.

While retroflexion is clearly an areal trait of South Asia, South Asian languages differ in the phonotactic restrictions that they place on retroflex consonants. This paper presents evidence that a large number of South Asian languages have developed a co-occurrence restriction on coronal obstruents that can be described as retroflex consonant harmony. In these languages, roots containing two non-adjacent coronal stops are primarily limited to those with two dentals (T…T) or two retroflexes (Ṭ…Ṭ), while those containing a combination of dental and retroflex stops are avoided (*T…Ṭ, *Ṭ…T). In addition, the paper demonstrates that the distribution of languages with and without retroflex consonant harmony is geographic in nature, not genetic. Retroflex consonant harmony is characteristic of most languages in the northern half of the subcontinent, regardless of whether they are Indo-Aryan, Dravidian or Munda. It is not characteristic of Indo-Aryan and Dravidian languages in the south. Thus, retroflex consonant harmony constitutes an areal feature within South Asia.

The study employs two kinds of evidence to support these conclusions: (i) statistical evidence of synchronic co-occurrence restrictions on coronal consonants in the vocabulary of a broad sample of languages, representing diverse genetic groups and geographic regions within South Asia, and (ii) historical-comparative evidence demonstrating that roots with retroflex consonant harmony can often be traced to disharmonic cognates in a parent language or in closely related languages or dialects.

The paper is organized as follows. Section 2 introduces a phonotactic restriction on retroflex consonants that has some antiquity in South Asia and contrasts it with a more innovative consonant harmony restriction. Taking Panjabi as a case study, section 3 illustrates the innovative pattern and its diachronic development. Sections 4-6 present evidence of parallel patterns and developments in a large number of Indo-Aryan, Dravidian and Munda languages. Section 7 then plots the distribution of languages with and without consonant harmony, demonstrating that each group defines a geographic region that cuts across genetic boundaries. Section 8 summarizes the typological properties of retroflex consonant harmony in South Asia and section 9 concludes the paper.

2. Two phonotactic patterns

While most South Asian languages have retroflex consonants, they differ with respect to the phonotactic restrictions that they place on those consonants. Two phonotactic patterns are of concern to us here. The first is a historical prohibition on word-initial retroflex consonants, which was characteristic of Old Indo-Aryan and Proto-Dravidian. The second is an innovative pattern of retroflex consonant harmony, which can be found in many contemporary Indo-Aryan, Dravidian and Munda languages. These are represented schematically in (1). Here and elsewhere T represents any dental stop and Ṭ represents any retroflex stop, abstracting away from laryngeal distinctions for voicing and aspiration.

(1) Two phonotactic patterns affecting the co-occurrence of dental (T) and retroflex (Ṭ) stops

a. No initial retroflexes		b. Retroflex consonant harmony	
T-T	T-Ṭ	T-T	*T-Ṭ
*Ṭ-T	*Ṭ-Ṭ	*Ṭ-T	Ṭ-Ṭ

Historically, Old Indo-Aryan and Proto-Dravidian avoided word-initial retroflex consonants (Schwarzschild 1973; Masica 1991; Zvelebil 1970; Subrahmanyam 1983; Krishnamurti 2003). As a result, roots containing two non-adjacent consonants were limited to just two of four possible configurations involving dental and retroflex stops: dental-dental (T-T) and dental-retroflex (T-Ṭ), but not retroflex-dental (*Ṭ-T) or retroflex-retroflex (*Ṭ-Ṭ). This phonotactic restriction is still maintained in most South Dravidian and southern Indo-Aryan languages, at least in native vocabulary. Elsewhere, most South Asian languages now admit word-initial retroflex stops, although they still avoid word-initial retroflex sonorants (nasals and liquids).

Many South Asian languages, which formerly had co-occurrence patterns like that in (1a), have developed consonant harmony systems like that in (1b). In this innovative pattern, co-occurring coronal stops must agree with respect to retroflexion (Ṭ-Ṭ) or non-retroflexion (T-T), while roots containing a combination of dental and retroflex stops are avoided (*Ṭ-T, *T-Ṭ). The pattern in (1b) has developed from (1a) largely through a process of long-distance retroflex assimilation, in which former dental-retroflex configurations have developed into retroflex-retroflex configurations (T-Ṭ → Ṭ-Ṭ). In order to better understand this development it will be useful to examine at least one case study in some detail before turning to the question of just how pervasive this development has been in South Asia.

3. Panjabi: A case study of retroflex consonant harmony

Panjabi is a New Indo-Aryan language spoken in the Punjab region in western India and adjacent parts of Pakistan. It is representative of those South Asian languages that have developed a pattern of retroflex consonant harmony like that in (1b), above. The consonants of Panjabi are listed in Table 1. Orthographic voiced aspirated stops are realized as their unaspirated counterparts with accompanying pitch contours on neighbouring vowels. In much of the literature, phonemic transcriptions mirror the orthography by using voiced aspirated characters to represent what is essentially a tonal contrast. This convention is retained here.

Table 1 Consonant phonemes of Panjabi (Malik 1995)

labial	dental	alveolar	retroflex	palatal	velar	glottal
p	t		ʈ	ʧ	k	
p^h	t^h		$ʈ^h$	$ʧ^h$	k	
b	d		ɖ	ʤ	g	
(b^h)	(d^h)		($ɖ^h$)	($ʤ^h$)	(g^h)	
		s		ʃ		h
m		n	ɳ			
		l	(ɭ)			
		r	ɽ			
w				j		

As shown in Table 1, Panjabi distinguishes retroflex stops and sonorants from their dental or alveolar counterparts. The retroflex lateral /ɭ/ is not phonemic in all dialects and is not distinguished in any of the data sources employed for the current study. Thus, it will not be considered further. The retroflex sonorants /ɳ/ and /ɽ/ are subject to the same phonotactic restriction that once applied to most retroflexes in OIA: they do not occur word-initially. However, retroflex stops, once avoided word-initially, now appear frequently in that position. Commenting on the historical development of Panjabi retroflex stops in word-initial position, Jain (1934: 89) points out a "tendency to cerebralise [make retroflex] a dental stop occurring in the vicinity of another cerebral [retroflex] stop", where "in the vicinity" means nearby but non-adjacent. The following sub-section examines synchronic co-occurrence restrictions on dental and retroflex stops in Panjabi. It reveals that the process of retroflex assimilation observed by Jain has produced a phonotactic pattern like that in (1b).

3.1. *Synchronic co-occurrence patterns in Panjabi*

We can measure the extent of consonant harmony in a language like Panjabi using simple statistical methods. Statistical methods are a convenient means of examining long-distance co-occurrence restrictions on consonants (Frisch et. al. 2004; Kawahara et. al. 2006; Coetzee and Pater 2008; Gallagher and Coon 2009). The method employed in much of the literature, and also adopted here, involves calculating the frequency with which consonants co-occur in the lexicon of a language. Using a lexical database, counts are made of non-adjacent C_1-C_2 pairs in the vocabulary of a language. These counts are referred to as *observed* values (O). Observed values are then used to derive *expected* values (E) for each pair. These are the values that would be expected if consonants co-occurred randomly in the data set. Observed-to-expected ratios (O/E) are then computed for each C_1-C_2 pair to determine whether some configurations occur more or less frequently than expected. An O/E ratio of 1.0 for a given C_1-C_2 pair indicates that there is no difference between the observed and expected frequencies for that pair. In other words, it occurs as expected and there is no restriction on it. An O/E ratio of more that 1.0 indicates that the C_1-C_2 pair occurs more frequently than expected and is favoured to some degree. An O/E ratio of less than 1.0 indicates that the C_1-C_2 pair occurs less frequently than expected and is avoided to

some degree. An O/E ratio of 0.0 indicates categorical absence, which is the strongest form of avoidance. Configurations that occur more frequently than expected are said to be *over-attested*. Those that occur less frequently than expected are said to be *under-attested*.

This statistical method was applied to data from Goswami's (2000) Panjabi-English dictionary. A count was made of dictionary headwords containing word-initial $C_1V(N)C_2$ sequences in which C_1 and C_2 are coronal stops (dental or retroflex) or retroflex sonorants, V is an intervening vowel and N is a homorganic nasal. Limiting the count to word-initial $C_1V(N)C_2$ sequences is a convenient way of approximating a count of root-internal sequences without attempting a morphological analysis of every word. This is possible because South Asian languages prefer suffixes to prefixes, lexical roots are often limited to one or two syllables, and homorganic nasal-stop sequences are common morpheme-internally. Retroflex sonorants were included in the count of C_2 in order to highlight their behaviour relative to stops. They do not occur in C_1 position. Sequences such as /C_1…ɳɖ/ were counted as instances of /C_1…ɖ/, not /C_1…ɳ/, because the place of articulation of the nasal is predictable on independent grounds in such cases. All counts of /C_1…ɳ/ reflect cases in which the nasal occurs without a following retroflex stop (e.g., /taɳaː/ 'trunk of a tree'). The results are shown in Table 2.

Table 2 Panjabi coronal stops and retroflex sonorants in #$C_1V(N)C_2$ sequences (n=233)

			C_2		
			T	Ṭ	Ṛ
C_1	T	O	70	5	70
		E	43.6	51.7	49.8
		O/E	**1.61**	**0.10**	**1.41**
	Ṭ	O	0	78	10
		E	26.4	31.3	30.2
		O/E	**0.00**	**2.49**	**0.33**

Table 2 should be read as follows. The y-axis (vertical) represents C_1 while the x-axis (horizontal) represents C_2. T represents the class of dental stops (/t, tʰ, d, dʰ/), Ṭ represents the class of retroflex stops (/ʈ, ʈʰ, ɖ, ɖʰ/) and Ṛ represents the class of retroflex sonorants (/ɳ, ɽ, ɽʰ/). Observed counts (O), expected counts (E), and observed/expected ratios (O/E) are listed for each possible C_1-C_2 configuration. Thus, there are 70 observed instances of a dental stop in C_1 position followed by another dental stop in C_2 position (T-T), whereas only 43.6 are expected. The O/E ratio for this configuration is 1.61, indicating that it occurs approximately one and a half times more frequently than expected. Similarly, there are 78 observed instances of a retroflex stop in C_1 position followed by another retroflex stop in C_2 position (Ṭ-Ṭ), whereas only 31.3 are expected. The O/E ratio for this configuration is 2.49, indicating that it occurs about two and a half times more frequently than expected. In contrast to this, there are only 5 observed instances a dental stop followed by a retroflex stop (T-Ṭ) even though as many as 51.7 are expected. The O/E ratio for this configuration is 0.10, indicating that it occurs far less frequently than expected. Configurations consisting of a retroflex stop followed by a dental stop (Ṭ-T) are categorically absent (O/E = 0.00). Thus, we can see that roots containing two coronal stops that agree in

retroflexion or non-retroflexion are overwhelmingly favoured (Ṭ-Ṭ, T-T) while those that disagree in retroflexion are avoided (*Ṭ-T, *T-Ṭ). Representative examples are listed in (2).

(2) Retroflex consonant harmony in Panjabi (data from Goswami 2000)
   a. Roots that agree in non-retroflexion (T-T)
      tatta: 'warm, hot'
      tand 'thread, strand'
      dand 'tooth'
      dudd^{h} 'milk'
   b. Roots that agree in retroflexion (Ṭ-Ṭ)
      ʈuʈɳa: 'to break'
      ʈiɖɖa: 'grass hopper'
      ʈunɖa: 'armless'
      ɖaɳɖa: 'stick'

Of the five disharmonic T-Ṭ sequences in Table 2, three are listed in Goswami (2000) as having variants that agree in retroflexion, as shown in (3a). The two remaining exceptions are *tatsamas* (i.e., unassimilated Sanskrit loans), as shown in (3b).

(3) Exceptions to retroflex consonant harmony in Panjabi (data from Goswami 2000)
   a. Variation between disharmonic and harmonic forms of the same root (T-Ṭ ~ Ṭ-Ṭ)
      toʈʈa: 'loss, deficiency' cf. ʈoʈa: 'deficiency, loss'
      toɖa: 'camel's young' cf. ʈoɖa: 'young camel'
      daɳɖ 'punishment' cf. ɖaɳɖ 'punishment'
   b. *Tatsamas*: unassimilated Sanskrit loanwords (T-Ṭ)
      taʈʈ 'coast, bank, shore' cf. Skt. taṭa- 'slope, shore'
      taʈast^{h} 'neutral' (lit. 'standing on a bank or shore') cf. Skt. taṭa- 'slope, shore'

Table 2 reveals another important fact. The distribution of retroflex sonorants differs from that of retroflex stops in C_2 position. Disharmonic T-Ṛ configurations are actually preferred (O/E = 1.41) over harmonic Ṭ-Ṛ configurations (O/E = 0.33). This is precisely the opposite of what we find with retroflex stops, where disharmonic T-Ṭ configurations are avoided and harmonic Ṭ-Ṭ configurations are favoured. This indicates that retroflex consonant harmony is a restriction specific to stops, not a restriction on coronal consonants in general. The phonotactic restriction on Panjabi stops illustrates the pattern of retroflex consonant harmony introduced in (1b). The following subsection examines the diachronic processes that have contributed to the development of this pattern.

3.2. *Historical-comparative evidence of consonant harmony*

The diachronic development of retroflex consonant harmony can be traced by comparing Panjabi roots with cognates in Old Indo-Aryan (OIA) Sanskrit and Middle Indo-Aryan (MIA) Prakrit. A comparison of this kind reveals that harmony in Panjabi is largely the product of assimilation that was long-distance and regressive. Original word-initial dental stops have become retroflex

whenever they were followed by a non-adjacent retroflex stop (T-Ṭ → Ṭ-Ṭ). The retroflex stop that triggered assimilation was typically part of a homorganic consonant cluster. In MIA this cluster was either a geminate (ṬṬ) or nasal-stop sequence (NṬ). These developments are illustrated in Table 3. Reference numbers in the right-most column refer to etymological groups in Turner's (1962–1966) *Comparative Dictionary of Indo-Aryan Languages* (CDIAL).

Table 3 Historical-comparative evidence of retroflex consonant harmony in Panjabi

		OIA	MIA	Panjabi	CDIAL
a.	'to be kind'	tuṣṭa-	tuṭṭʰa-	ṭuṭṭʰnaː	5895
	'fallen'	dʰvasta- > *dʰvaṣṭa-	*dʰaṭṭʰa-	ɖʰaːṭʰ	6896
b.	'to open'	tardati	taḍḍaï	ṭaḍḍṇaː	5721
	'frog'	dardura-	daddura, ḍeḍḍura	ɖaḍḍuː	6198
	'strong'	dṛɖʰa-, daːrɖʰja-	diḍɖʰa-, daɖʰa-	ɖaːɖɖʰaː	6508
	'lamp stand'	diːpavarti	—	ɖiũːṭ	6354
c.	'to break'	truṭjati	tuṭṭaï, ṭuṭṭaï	ṭuṭṭṇaː	6065
	'screen'	—	taṭṭiː-, ṭaṭṭiː-	ṭaṭṭiː	5990
	'grasshopper'	—	tiḍḍa-	ṭiḍḍaː	6024
	'plug'	—	-daṭṭa-	ɖaṭṭaː	6618
d.	'stick'	daṇḍa-	daṁḍa, ɖaṁḍa-[1]	ɖaṇḍaː	6128
	'vessel for curd'	dadʰibʰaːṇḍa	—	ɖahiṇḍi	6149

In Table 3, the examples in (a)–(c) illustrate assimilation of an initial dental stop to what was most likely a following geminate retroflex stop. In most cases the geminate is attested in MIA and still preserved in Panjabi. Stem-internal geminates were rare in OIA. They developed in MIA from various OIA consonant clusters. The geminates in (a) are reflexes of OIA /ṣṭ/ clusters; those in (b) derive from OIA /rC/ clusters in which C was a coronal stop; and those in (c) are of other (e.g., -ṭj- > -ṭṭ-) or uncertain origin. The examples in (d) illustrate cases where the retroflex stop that triggered assimilation was part of a homorganic nasal-stop cluster. Notice that some MIA cognates show variation between harmonic and dis-harmonic forms of the same root (e.g., /taṭṭiː-/ ~ /ṭaṭṭiː-/ 'screen'). This suggests that retroflex consonant harmony was already at work in some dialects during the MIA period (cf. Schwarzschild 1973: 484).

In South Asian languages, single intervocalic retroflex stops are commonly subject to lenition. Lenition of intervocalic OIA and MIA retroflex stops has produced sonorant retroflex flaps in most New Indo-Aryan (NIA) languages, either as allophones of the stops or (via subsequent developments) as independent retroflex phonemes, /ɽ/ and /ɽʰ/. With few exceptions, these flaps do not trigger retroflex consonant harmony. Likewise, retroflex nasals that occur outside of homorganic nasal-stop clusters typically do not trigger retroflex consonant harmony. Representative examples from Panjabi are listed in Table 4.

Table 4 No harmony between stops and retroflex sonorants in Panjabi

		OIA	MIA	Panjabi	CDIAL
a.	'palm tree'	*ta:ɖa-	ta:ɖa-	ta:ɽ	5750
	'to break'	tro:ṭajati	to:ɖaï	to:ɽna:	6079
	'robbery'	$d^ha:ṭi:$	$d^ha:ɖi:-$	$d^ha:ɽa:$	6772
	'beard'	$da:ɖ^hika:$	$da:ɖ^hia:$	$da:ɽ^hi:$	6250
b.	'tune'	ta:na	ta:ɳa-	ta:ɳ	5761
	'teat, udder'	stana	$t^haɳa-$	$t^haɳ$	13666
	'gift, charity'	da:na	da:ɳa-	da:ɳ	6265
	'incense burning'	$d^hu:pana$	$d^hu:vaɳa-$	$d^hu:ɳi:$	6848

The diachronic evidence confirms that long-distance assimilation has contributed to the phonotactic co-occurrence pattern observed in Table 2. Panjabi is not unique in this respect. Examples of long-distance retroflex assimilation have been noted in the historical phonology of other South Asian languages. The rest of this paper explores the full extent of retroflex consonant harmony in South Asia.

4. Indo-Aryan

In order to determine the extent of retroflex consonant harmony in South Asia, the statistical method illustrated in the Panjabi case study was applied to lexical data from a broad sample of languages representing different genetic groups and geographic regions. In the case of Indo-Aryan, two languages were chosen to represent each geographic zone, following the classification in Lewis (2009): Indus Kohistani and Sindhi for the Northwestern zone; Kumauni and Nepali for the Northern zone; Panjabi and Hindi for the Central zone; Bangla and Oriya for the Eastern zone; Marathi and Konkani for the Southern zone; and Sinhala and Dhivehi for the Sinhalese-Maldivian group.[2] Data for each language was obtained from a dictionary or vocabulary list. Counts were then made of word-initial $\#C_1V(N)C_2$ sequences in which C_1 and C_2 are coronal stops or retroflex sonorants.

For ease of readability the results are presented in Table 5 following a convention introduced by Pozdniakov and Segerer (2007). Rather than presenting O/E ratios, Pozdniakov and Segerer measure the discrepancy between observed and expected values and express it as a positive or negative percentage. For example, if a particular C_1-C_2 configuration has an O/E value of 1.25 then we might say that it is over-attested by +25%. Similarly, if a C_1-C_2 configuration has an O/E value of 0.75 then we might say that it is under-attested by −25%. The results are then presented schematically as follows: (i) a discrepancy whose absolute value is less than 25% is considered non-significant and is not noted (i.e., the cell in the table is left empty); (ii) a discrepancy whose absolute value is between 25% and 50% is represented by a single "+" or "−" sign; (iii) a discrepancy whose absolute value is greater than 50% is noted by a double "+ +" or "− −" sign.[3] To further aid readability, cells with under-attested values are shaded grey.

Table 5 Coronal stops and retroflex sonorants in #$C_1V(N)C_2$ sequences in eight Indo-Aryan languages of the Northwestern, Northern, Central and Eastern zones

Indus Kohistani (Zoller 2005; n=150)

	T	Ṭ	Ṛ
T	+ +	− −	+
Ṭ	− −	+ +	− −

Sindhi (Turner 1969; n=106)

	T	Ṭ	Ṛ
T	+	− −	
Ṭ	− −	+ +	− −

Kumauni (Van Riezen, p.c.; n=54)

	T	Ṭ	Ṛ
T	+	− −	
Ṭ	− −	+ +	

Nepali (Turner 1931; n=597)

	T	Ṭ	Ṛ
T	+ +	− −	− −
Ṭ	− −	+ +	+ +

Panjabi (Goswami 2000; n=233)

	T	Ṭ	Ṛ
T	+ +	− −	+
Ṭ	− −	+ +	− −

Hindi (McGregor 1993; n=777)

	T	Ṭ	Ṛ
T	+	− −	
Ṭ	− −	+ +	

Bangla (Biswas 2000; n=357)

	T	Ṭ	Ṛ
T		− −	
Ṭ	− −	+ +	−

Oriya (Turner 1969; n=165)

	T	Ṭ	Ṛ
T	+		
Ṭ	− −	+ +	

Two observations can be made about the results in Table 5. First, with the exception of Oriya, every language in the table exhibits a pattern of retroflex consonant harmony between stops. For example, in Indus Kohistani, Ṭ-T and T-Ṭ configurations, which disagree in retroflexion, are both represented by double "− −" signs, indicating that they are under-attested by more than −50%. At the same time, Ṭ-Ṭ configurations that agree in retroflexion, and T-T configurations that agree in non-retroflexion, are both represented by double "+ +" signs, indicating that they are over-attested by more than +50%. Thus, roots containing two coronal stops that agree in retroflexion or non-retroflexion are clearly favoured over those that disagree. The same trend holds for every other language except Oriya. In Oriya, disharmonic T-Ṭ configurations are represented by an empty cell, indicating that they occur more or less as expected. In fact, T-Ṭ configurations are statistically under-attested in Oriya but only by −19% (O/E = 0.81). Thus, while harmonic Ṭ-Ṭ configurations occur more frequently than expected in Oriya, disharmonic T-Ṭ configurations are not strongly avoided.

Second, with the exception of Nepali, no language exhibits a harmony pattern between stops and sonorants. In most cases, retroflex sonorants co-occur with stops as expected or disharmonic T-Ṛ sequences are favoured and harmonic Ṭ-Ṛ sequences are avoided. The only exception to this rule is Nepali, which appears to favour harmonic Ṭ-Ṛ sequences. This anomaly can be explained by the fact that Ṭ-Ṛ configurations in Nepali reflect phonemic /Ṭ-Ṭ/ configurations. As a result, they are cognate with Ṭ-Ṭ configurations in the other NIA languages, not with T-Ṛ or Ṭ-Ṛ configurations in those languages. Consider the examples in Table 6. In Panjabi, the retroflex

flaps /ɽ/ and /ɽʰ/ are reflexes of intervocalic /-ɖ-/ and /-ɖʰ-/ in MIA. In Nepali, however, these segments have merged with /r/, as illustrated in Table 6 (a). MIA /ɖ/ and /ɖʰ/ are preserved as stops whenever they occurred in geminate or homorganic nasal-stop clusters. This is true of both Panjabi and Nepali. However, in the case of Nepali, these stops are often realized allophonically as [ɽ], as illustrated in Table 6 (b). Whenever the retroflex stop in question was preceded by a homorganic nasal, Turner (1931) records free variation between [Ṽɽ] and [Vɳ]. Thus, all Ṭ-Ṛ configurations in Nepali reflect underlying /Ṭ-Ṭ/ configurations. As a result, they exhibit the same assimilatory tendency as Ṭ-Ṭ configurations in other NIA languages.

Table 6 Reflexes of MIA /ɖ/ and /ɖʰ/ in Panjabi and Nepali

	MIA	Panjabi		Nepali		CDIAL
a.	t…ɖ	taːɽ	'palm tree'	taːr	'palm (tree)'	5750
	d…ɖʰ	daːɽʰiː	'beard'	daːri	'beard'	6250
b.	*t…ɖɖʰ	ʈe(ɖ)ɖʰaː[4]	'crooked'	ʈeɽo	'slanting'	6071
	d…ɳɖ	ɖaɳɖiː	'a thin stick'	ɖãːɽi ~ ɖaːɳi	'pole, staff'	6128
	*d…ɳɖ	ɖuɳɖ	'headless body'	ɖũɽo ~ ɖuɳo	'maimed'	5560

The Northwestern languages have some unique properties that deserve mention. For example, Sindhi has a series of implosive stops. The opposition between dental and retroflex is neutralized in the implosive series, where only alveolar /ɗ/ occurs. Implosive /ɗ/ does not participate in consonant harmony. It co-occurs with dental and retroflex plosives alike (e.g., /taɗo/ 'pit', /ʈuɗo/ 'crippled'). Like most Northwestern languages of the Dardic group, Indus Kohistani has retroflex sibilant affricates and fricatives in addition to stops. Retroflex consonant harmony applies to co-occurring stops (e.g., /ɖíːʈʰi/ 'span of hand'), and co-occurring sibilants (e.g., /ʂìʂ/ 'head'), but not between stops and sibilants (e.g., /dùːʂ/ 'sin', /sìːʈʰ/ 'rich'). A similar pattern occurs in Kalasha and other Dardic languages (Arsenault and Kochetov 2011; Arsenault 2012).

The evidence examined thus far indicates that retroflex consonant harmony is most pervasive in the Northwest, Northern and Central zones and weakest in the Eastern zone, where it is marginal at best in Oriya. Further evidence indicates that it is absent altogether in languages of the Southern and Sinhalese-Maldivian zones. Like OIA, languages in the Southern and Sinhalese-Maldivian zones generally avoid word-initial retroflex consonants and preserve disharmonic T-Ṭ configurations. This is evident in Table 7, which lists observed counts for co-occurring coronal stops and retroflex sonorants in a sample of languages from those zones. Observed counts are reported here because O/E ratios can be misleading in languages that avoid word-initial retroflexes. The near absence of word-initial retroflex consonants leads to extremely low expected frequencies for configurations involving initial retroflexes. Low expected values, in turn, yield exaggerated and unreliable O/E ratios. For example, the expected frequency of Ṭ-Ṛ configurations in Konkani is just 0.3. With an expected value this low, a single observed instance leads to an O/E ratio of 3.33. This is misleading because it suggests that harmonic Ṭ-Ṛ sequences are highly favoured when in fact there is only one example, as compared to 37 examples of disharmonic T-Ṛ sequences. For this reason, O/E ratios are not computed for languages that

avoid word-initial retroflexes. Instead, simple observed values are allowed to speak for themselves.[5]

Table 7 Observed counts for coronal stops and retroflex sonorants in #C_1V(N)C_2 sequences in NIA languages of the Southern and Sinhalese-Maldivian zones

Marathi (Molesworth 1857; n=1833)

	T	**Ṭ**	**Ṛ**
T	463	769	316
Ṭ	2	55	228

Konkani (Maffei 1883; n=127)

	T	**Ṭ**	**Ṛ**
T	41	48	37
Ṭ	0	0	1

Sinhala (Turner 1969; n=90)

	T	**Ṭ**	**Ṛ**
T	43	27	20
Ṭ	0	0	0

Dhivehi (Reynolds 2003; n=106)

	T	**Ṭ**	**Ṛ**
T	37	21	44
Ṭ	0	4	0

In each of the languages in Table 7, disharmonic T-Ṭ and T-Ṛ configurations are frequent while harmonic Ṭ-Ṭ and Ṭ-Ṛ configurations are infrequent or absent altogether. Those Ṭ-Ṭ and Ṭ-Ṛ configurations that do occur are often the product of independent factors such as reduplication, onomatopoeia or loanword adaptation. For example, many Ṭ-Ṭ and Ṭ-Ṛ configurations in Molesworth's (1857) Marathi dictionary are onomatopoeic expressions with reduplication (e.g., /ʈʰuʈʰaː/ 'the rattle of musket-firing', /ɖʰaɳaɖʰaɳa/ 'imit. of clanging'). As a general rule, the languages in Table 7 have preserved inherited T-Ṭ configurations.

The conclusions regarding retroflex consonant harmony in Indo-Aryan are supported by historical-comparative evidence. Consider the data in Table 8. The examples in part (a) represent C_1-C_2 configurations in which C_2 was a retroflex stop that was either geminate or part of a homorganic nasal-stop cluster in MIA. Attested MIA cognates for these roots are typically disharmonic although some have harmonic variants (T-Ṭ or T-Ṭ~Ṭ-Ṭ). In every case, Marathi (representing the Southern zone) exclusively preserves disharmonic forms (T-Ṭ). Bangla (Eastern), Hindi (Central) and Nepali (Northern) all favour harmonic forms, with disharmonic variants preserved in a few cases (Ṭ-Ṭ or T-Ṭ~Ṭ-Ṭ). The examples in part (b) represent configurations in which C_2 was an intervocalic retroflex stop in MIA. These roots exhibit a different trend. Attested MIA cognates are exclusively disharmonic (T-Ṭ) and disharmonic forms are favoured in all of the languages, with harmonic variants in isolated cases (T-Ṛ or T-Ṛ~Ṭ-Ṛ).

Table 8 Historical-comparative evidence of retroflex consonant harmony in Indo-Aryan

		MIA	Marathi (S)	Bangla (E)	Hindi (C)	Nepali (N)	CDIAL
a.	'pony'	—	taʈʈuː, taʈuː	ʈaːʈu	ʈaʈʈuː	ʈaʈu	5440
	'woven work'	taʈʈiː, ʈaʈʈiː	taːʈiː, taːʈ	ʈaːʈi, ʈaːt	ʈaːʈʈiː, ʈaːt	taːʈ, ʈaːt	5990
	'to be broken'	tuʈʈaï, ʈuʈʈaï	tuʈɳẽ	ʈuʈaː	ʈuːʈnaː	ʈuʈnu	6065
	'seeing, sight'	diʈʈʰi	diːʈʰ	—	diːʈʰ, ɖiːʈʰ	ɖiʈʰ	6520
	'to threaten'	-daʈʈa-,*daṁʈ-	daːʈɳẽ	ɖãːʈaː	ɖaːʈnaː, ɖãːʈnaː	ɖãːʈnu	6618
	'mouth, spout'	tuṁɖa, toṁɖa	tũɖ, tõɖ	ʈũʈi	ʈõʈ, tõʈiː	ʈũʈo, ʈuʈo	5853
	'stick, club'	daṁɖa, ɖaṁɖa	dãːɖaː	daɳɖa, ɖaːɳɖaː	daɳɖaː, ɖaɳɖaː	ɖãːʈi	6128
	'cold'	—	tʰaɳɖaː	ʈʰaːɳɖaː	ʈʰaɳɖʰaː	ʈʰaɳɖaː	13676
b.	'palm (tree)'	taːɖa	taːɖ	taːʈ	taːʈ	taːr	5750
	'large tooth'	daːɖʰaː	daːɖʰ	daːʈ, ɖaːʈ	daːʈʰ, ɖaːʈʰ	daːro	6250
	'chin, beard'	daːɖʰiaː	daːɖʰiː	daːʈi	daːʈʰiː, ɖaːʈʰiː	daːri	6250
	'assault'	dʰaːɖiː	dʰaːɖiː, dʰaːɖ	—	dʰaːʈ	—	6772

In summary, retroflex consonant harmony between stops is pervasive in NIA languages of the Northwestern, Northern and Central zones. In the Eastern zone it occurs in Bangla but only marginally in Oriya. It does not apply to NIA languages of the Southern or Sinhalese-Maldivian zones. In the following section we turn our attention to Dravidian.

5. Dravidian

Retroflex consonant harmony has been reported in a number of Dravidian languages. The North Dravidian subgroup consists of only three major languages: Malto, Kurux and Brahui. Of these, Brahui is geographically isolated in Pakistan and does not exhibit evidence of consonant harmony. However, retroflex consonant harmony has been reported in Malto (Mahapatra 1979) and Kurux (Pfeiffer 1972), both of which are spoken in eastern India. Moreover, Burrow & Bhattacharya (1963) report retroflex consonant harmony in Kuvi, a language of the South-Central group, which is also spoken in eastern India. They point out that the pattern in Kuvi is characteristic of "most of the Dravidian languages of this area" (1963: 240). The languages they name include four South-Central languages, Kuvi, Kui, Pengo and Konda, and two Central Dravidian languages, Parji and Gadaba (cf. Subrahmanyam 1983).

A survey of data from these languages confirms the presence of consonant harmony in each one, with the exception of Gadaba. Table 9 shows the co-occurrence of coronal stops and retroflex sonorants in $\#C_1V(N)C_2$ sequences from eight Dravidian languages of the Northern, South-Central and Central groups. Again, for ease of readability the results are presented following the convention of Pozdniakov and Segerer (2007).

Table 9 Coronal stops and retroflex sonorants in #C_1V(N)C_2 sequences in eight Dravidian languages of the Northern, South-Central and Central groups

Kurux (Grignard 1986 [1924]; n=286)

	T	Ṭ	Ṛ
T	+ +	− −	
Ṭ	− −	+ +	

Malto (Mahapatra 1987; n=137)

	T	Ṭ	Ṛ
T	+ +	− −	+ +
Ṭ	− −	+ +	− −

Kui (Winfield 1929; Burrow & Bhattacharya 1961; Burrow & Emeneau 1984; n=42)

	T	Ṭ	Ṛ
T	+ +	− −	+ +
Ṭ	− −	+	−

Pengo (Burrow & Bhattacharya 1970; n=77)

	T	Ṭ	Ṛ
T	+ +	− −	
Ṭ	− −	+ +	

Kuvi (Israel 1979; n=83)

	T	Ṭ	Ṛ
T	+ +	− −	+ +
Ṭ	− −	+	−

Konda (Krishnamurti 1969; n=37)

	T	Ṭ	Ṛ
T	+ +	− −	+ +
Ṭ	− −	+ +	−

Gondi (Penny et. al. 2005; n=92)

	T	Ṭ	Ṛ
T	+ +	−	+
Ṭ	− −	+ +	− −

Parji (Burrow & Bhattacharya 1953; n=86)

	T	Ṭ	Ṛ
T	+ +	− −	+ +
Ṭ	− −	+ +	− −

The Dravidian languages in Table 9 bear a strong resemblance to the Indo-Aryan languages in Table 5. To begin with, all languages in Table 9 exhibit a pattern of retroflex consonant harmony between coronal stops. Roots containing stops that agree in retroflexion (Ṭ-Ṭ) are highly favoured (+ or + +) while those containing stops that disagree in retroflexion (Ṭ-T, T-Ṭ) are avoided (− or − −). Gondi appears to preserve a number of T-Ṭ forms alongside innovative Ṭ-Ṭ forms. The mixed pattern may be the result of dialect variation. Many Gondi roots in Penny et. al. (2005) are listed as having harmonic and disharmonic variants (e.g., /teɳḍ- ~ ṭeɳḍ-/ 'to take out'). Secondly, roots containing retroflex sonorants show just the opposite trend. Disharmonic T-Ṛ configurations occur at or above expected frequencies (empty cell, + or ++), while harmonic Ṭ-Ṛ configurations occur at or below expected frequencies (empty cell, − or − −). No language shows a preference for Ṭ-Ṛ configurations. Thus, every language in Table 9 exhibits a pattern of retroflex consonant harmony between stops but not, as a rule, between stops and sonorants.

Retroflex consonant harmony holds for most Dravidian languages of the South-Central group, but not for Telugu, the only major literary language of that group. It clearly extends to Central Dravidian Parji, but it does not extend to any other Central Dravidian language, including Gadaba (contra Burrow and Bhattacharya [1963]). Moreover, retroflex consonant harmony does not apply to any language of the South Dravidian group. All Dravidian languages that lack retroflex consonant harmony exhibit a phonotactic pattern like the one attributed to Proto-Dravidian: retroflex consonants are avoided word-initially. Table 10 shows observed counts for

coronal stops and retroflex sonorants in #C_1V(N)C_2 sequences from three Dravidian languages with this pattern: Telugu (South-Central), Gadaba (Central) and Tamil (South).

Table 10 Observed counts for coronal stops and retroflex sonorants in #C_1V(N)C_2 sequences from three Dravidian languages without consonant harmony

Telugu (Gwynn 1991; n=519)

	T	**Ṭ**	**Ṛ**
T	176	293	43
Ṭ	1	5	1

Gadaba (Bhaskararao 1980; n=38)

	T	**Ṭ**	**Ṛ**
T	4	25	8
Ṭ	0	1	0

Tamil (Fabricius 1972; n=612)

	T	**Ṭ**	**Ṛ**
T	124	254	230
Ṭ	0	2	2

The counts in Table 10 confirm that these languages do not exhibit retroflex consonant harmony. In each case, disharmonic T-Ṭ configurations have the highest observed counts while harmonic Ṭ-Ṭ configurations are among the lowest. Recall that Burrow and Bhattacharya (1963) list Gadaba among the Dravidian languages with retroflex consonant harmony. The results in Table 10 do not bear this out. The single example cited by those authors is /ʈeʈp-/ 'to raise', which appears to be an exception.

Historical-comparative data confirms that retroflex consonant harmony has applied in most North (N) and South-Central (SC) languages and in Central (C) Dravidian Parji. Examples are listed in Tables 11, 12 and 13. In each table, harmonic word forms from the respective language group are identified with disharmonic cognates from Telugu, representing a conservative variety of South-Central Dravidian, and Tamil, representing South (S) Dravidian. In the few instances where a cognate from Tamil was not available, a cognate from Malayalam (Ma.) has been supplied instead. Reference numbers in the right-most column of each table refer to etymological groups in Burrow and Emeneau's (1984) revised *Dravidian Etymological Dictionary* (DEDR).

Table 11 Evidence of retroflex consonant harmony in North Dravidian

	Tamil (S)	**Telugu (SC)**	**Kurux (N)**	**Malto (N)**	**DEDR**
'to grope'	taʈavu	taɖavu	–	ʈoʈr-	3025[6]
'to hinder'	taʈu, taɳʈu	taʈa:jinʧu	ʈaɳɖna:	–	3031
'side'	taʈʈu	taʈʈu	ʈo:ʈ	–	3040
'strength'	tiɳ, tiɳʈi	daɳɖi (?)	ʈinɖi:	–	3222
'to smear, wipe'	tuʈai	tuɖuʧu	ʈu:ɽna:, ʈuɖɖjas	ʈuɖ	3301
'big(-bellied)'	toʈʈa	doɖɖa, doɖɖu	ɖonɖa:	ɖuɖwa	3491

Table 12 Evidence of retroflex consonant harmony in South-Central Dravidian

	Tamil (S)	Telugu (SC)	Gondi (SC)	Konda (SC)	Kuvi (SC)	DEDR
'many, thick'	taṯam	daṯṯamu	da:ṯ, ḍa:ṯ	ḍaṯam	–	3020
'upper arm'	taṉṯa (Ma.)	daṉḍa	danḍ, ḍanḍ	ḍaṉḍa	ḍaṉḍe	3048
'kind of tree'	ta:ṉṯi	tã:ḍi, ta:ṇḍra	–	ṯa:ṇḍi	ṯa:ṇḍi	3198
'stout stick'	tuṯṯu-kkaṯṯai	duḍḍu	ḍuḍu	ḍudu	ḍudu	3304
'backyard'	toṯṯi	doḍḍi	ḍoḍḍi, ḍiḍḍi	ḍodi	ḍodi	3485
'beehive'	ti:n-toṯai	te:ne-teṯṯu	teṯṯe, ṯeṯṯe	te:ne-ṯaṯa	ṯaṯṯi, ṯaṯṯo	3490
'big'	toṯṯa	doḍḍa, doḍḍu	–	–	ḍoṯa	3491
'creeper'	toṉṯai	doṇḍa	ṯonḍa, ṯonḍri	ḍoṇḍa	–	3499
'garden'	to:ṯṯam	tõ:ṯa	–	ṯo:ṇṯa	ṯo:ṯa	3549
'companion'	to:ɻaṉ	to:ḍu	-to:ɽo	to:ɽu	to:ɽu	3563

Table 13 Evidence of retroflex consonant harmony in Central Dravidian Parji

	Tamil (S)	Telugu (SC)	Gadaba (C)	Parji (C)	DEDR
'upper arm'	taṉṯa (Ma.)	daṇḍa	–	ḍanḍa	3048
'to pull'	–	–	tinḍ-	ṯanḍ-	3052
'to touch'	toṯu, toṯṯ-	toṯṯu	–	ṯoḍ-, ṯoṯṯ-	3480
'creeper'	toṉṯai	doṇḍa	–	ṯunḍa	3499
'labourer'	to:ṯṯi	to:ṯi	–	to:ṯa, ṯo:ṯal	3546
'rake/hook'	to:ṯṯi	dõ:ṯi	–	ṯo:ṯal	3547
'to draw water'	to:ṉṯu	tõ:ḍu	to:ṇ-, to:nḍ-	ṯõ:ḍ-	3549
'to show'	to:ṯṯu (to:ṯṯi-)	–	to:ṯp-	ṯo:ṯip-	3566

Examining the correspondences in Tables 11–13, we see that harmonic Ṭ-Ṭ forms in languages with consonant harmony correspond to disharmonic T-Ṭ forms in those without. This affirms that harmony is the product of regressive long-distance assimilation between stops (T-Ṭ → Ṭ-Ṭ). With few exceptions, the retroflex stop in C_2 position that triggered assimilation corresponds to a geminate or homorganic nasal-stop cluster in conservative languages, such as Tamil and Telugu. In addition to dental and retroflex consonants, Proto-Dravidian also had a series of apical alveolar stops and nasals, represented here as /ṯ/ and /ṉ/. These are preserved in a few South Dravidian languages, including dialects of Tamil. Elsewhere the alveolar series has merged with dentals or retroflexes. In cases where the alveolars have developed into retroflex stops, they have triggered retroflex assimilation in any preceding dental stops. This is evident in examples such as Kuvi and Konda /ṯa:ṇḍi/ 'kind of tree', which corresponds to Tamil /ta:ṉṯi/ in Table 12, and Parji /ṯo:ṯip-/ 'to show', which corresponds to Tamil /to:ṯṯi-/ in Table 13. Finally, in Table 12, the retroflex approximant /ɻ/ in roots such as Tamil /to:ɻaṉ/ corresponds to a retroflex flap /ɽ/ in Kuvi and Konda /to:ɽu/ 'companion' and Gondi /-to:ɽo/. These sonorants have not triggered retroflex harmony in preceding stops.

In summary, retroflex consonant harmony is characteristic of most North Dravidian languages, including Malto and Kurux, but not Brahui; most South-Central Dravidian languages, including Kui, Kuvi, Pengo, Konda and (to a lesser extent) Gondi, but not Telugu; and at least

one Central Dravidian language: Parji. It does not appear to affect any other Central Dravidian language, including Gadaba, nor does it appear to affect any language of the South Dravidian group, which constitutes the majority of Dravidian languages.

6. Munda

The Munda family consists of a small set of minority languages concentrated in eastern India. A survey of Munda languages reveals that most of them exhibit the same pattern of retroflex consonant harmony found in Dravidian and Indo-Aryan languages. Table 14 shows the co-occurrence of coronal stops and retroflex sonorants in four North Munda languages and Table 15 does the same for six South Munda languages. Results in both tables are based on counts of #C_1V(N)C_2 sequences in which C_1 and C_2 are coronal stops or retroflex sonorants.

Table 14 Coronal stops and retroflex sonorants in #C_1V(N)C_2 sequences in four North Munda languages

Mundari (Bhaduri 1983 [1931]; n=198)

	T	Ṭ	Ṛ
T	+ +	− −	+
Ṭ	− −	+ +	−

Santali (Bodding 1929−1936; n=1315)

	T	Ṭ	Ṛ
T	+ +	− −	+
Ṭ	− −	+ +	−

Korwa (George & Joseph, p.c.; n=49)

	T	Ṭ	Ṛ
T	+ +	− −	
Ṭ	− −	+ +	

Ho (Deeney 1978; n=178)

	T	Ṭ	Ṛ
T	+ +	−	+
Ṭ	− −	+ +	− −

Table 15 Coronal stops and retroflex sonorants in #C_1V(N)C_2 sequences in six South Munda languages

Kharia (Peterson 2009; n=128)

	T	Ṭ	Ṛ
T	+ +	− −	
Ṭ	− −	+ +	

Juang (Donegan & Stampe 2004; n=68)

	T	Ṭ	Ṛ
T	+ +	− −	
Ṭ	− −	+ +	

Remo (Donegan & Stampe 2004; n=53)

	T	Ṭ	Ṛ
T	+ +	− −	
Ṭ	− −	+ +	

Gorum (Donegan & Stampe 2004; n=55)

	T	Ṭ	Ṛ
T	+ +	−	
Ṭ	− −	+	

Gtaʔ (Donegan & Stampe 2004; n=25)

	T	Ṭ	Ṛ
T	+ +	− −	+ +
Ṭ	− −	+	− −

Gutob (Donegan & Stampe 2004; n=32)

	T	Ṭ	Ṛ
T	+ +	− −	
Ṭ	− −	+	

The pattern exhibited by Munda languages in Tables 14 and 15 is the same as that exhibited by Dravidian and Indo-Aryan languages with retroflex consonant harmony. In every case, roots containing two stops that agree in retroflexion or non-retroflexion (Ṭ-Ṭ, T-T) are over-attested (+ or + +) while those that disagree in retroflexion (T-Ṭ, Ṭ-T) are under-attested (− or − −). The co-occurrence of stops with retroflex sonorants shows just the opposite trend. Disharmonic T-Ṛ sequences occur either at or above expected frequencies (empty cell, + or + +) while harmonic Ṭ-Ṛ sequences occur either at or below expected frequencies (empty cell, − or − −). Thus, every Munda language surveyed exhibits a pattern of retroflex consonant harmony between coronal stops but not between stops and sonorants.

Disharmonic T-Ṭ sequences are not under-attested to the same degree in every Munda language. They are only partly under-attested in Ho (Table 14) and Gorum (Table 15), as indicated by the single "−" sign. These languages bear a resemblance to Dravidian Gondi, in which many T-Ṭ forms are preserved alongside harmonic Ṭ-Ṭ forms (cf. Table 9). Variation between T-Ṭ and Ṭ-Ṭ is explicitly recorded in Deeney's (1978) Ho dictionary. A few examples are listed in Table 16 along with cognates from Mundari and Santali.

Table 16 Evidence of retroflex consonant harmony in North Munda

		Ho	Mundari	Santali
a.	'blunted arrow-head'	tuʈi, ʈuʈi	ʈuʈi, ʈoʈe	ʈuʈi
	'stump of tree'	duʈu, ɖuʈu	ɖuʈu	—
	'stick'	danɖa', ɖanɖa'	ɖaːnʈaː, ɖaːnɖaː	ɖanʈa
	'virgin'	dinɖa, ɖinɖa,	ɖinɖaː	—
b.	'to spread the legs'	tanɖa	ʈanɖaː	ʈanɖa
	'to lean'	tenɖer	ʈenɖer	ʈenɖar
	'to tie in a knot'	tonɖom	ʈonɖom	—
	'tooth'	daʈa	ɖaːʈaː	ɖaʈa
	'claw of a crab'	daɽo	ɖaɖo	ɖaʈɔm
c.	'restless, active'	toɽoboɽi, toɽoboɽo	taɽbaɽiaː	taɽbəɽiə
	'purse'	—	toɽaː	toɽa
	'to sting'	tuː	tuɽ	toɽ
	'squirrel'	tuː	tuɽa	toɽ

Wherever disharmonic T-Ṭ forms are attested for Ho in Table 16 they correspond to harmonic Ṭ-Ṭ forms in Mundari and Santali. The examples in section (a) of Table 16 show dialectal variation between harmonic and disharmonic forms in Ho itself and demonstrate that harmonic forms are always preferred in the other languages. Those in section (b) demonstrate the same preference even when variation is not reported for Ho. When retroflex flaps occur in C_2 position all of the languages prefer disharmonic T-Ṛ forms, as shown in section (c) (cf. Ho /daɽo/ in [b]). The dialectal variation within Ho and the comparison of data across closely related languages indicate that long-distance assimilation has contributed to the development of retroflex consonant harmony in Munda, just as it has in Dravidian and Indo-Aryan.

7. Retroflex consonant harmony as an areal feature

The evidence surveyed in the preceding sections indicates that retroflex consonant harmony is widespread in South Asia and cuts across genetic boundaries. Having established this, we can now demonstrate that the geographic distribution of languages with and without retroflex consonant harmony follows a clear trend. Setting aside Tibeto-Burman, for which we have no evidence of retroflex harmony, we can say that languages with retroflex consonant harmony are concentrated in the northern half of the Indian sub-continent while those that lack retroflex harmony, and avoid word-initial retroflex consonants, are concentrated in the south. This is demonstrated in Figure 1, which shows the approximate location of languages with and without retroflex consonant harmony in South Asia. Languages that retain a large number of disharmonic T-Ṭ forms alongside innovative Ṭ-Ṭ forms are classified as having *partial harmony*. In the interest of space, numbers are used in lieu of language names in Figure 1. An index of languages and their corresponding numbers is provided in Table 17. In addition to the languages discussed above, the list includes Kalasha, Malayalam and Kannada. Retroflex consonant harmony has been independently documented in Kalasha (Arsenault and Kochetov 2011; Arsenault 2012). Malayalam and Kannada are included to round out the picture of South Dravidian. The absence of retroflex consonant harmony in these and other South Dravidian languages is abundantly clear from historical-comparative data in Burrow and Emeneau (1984).

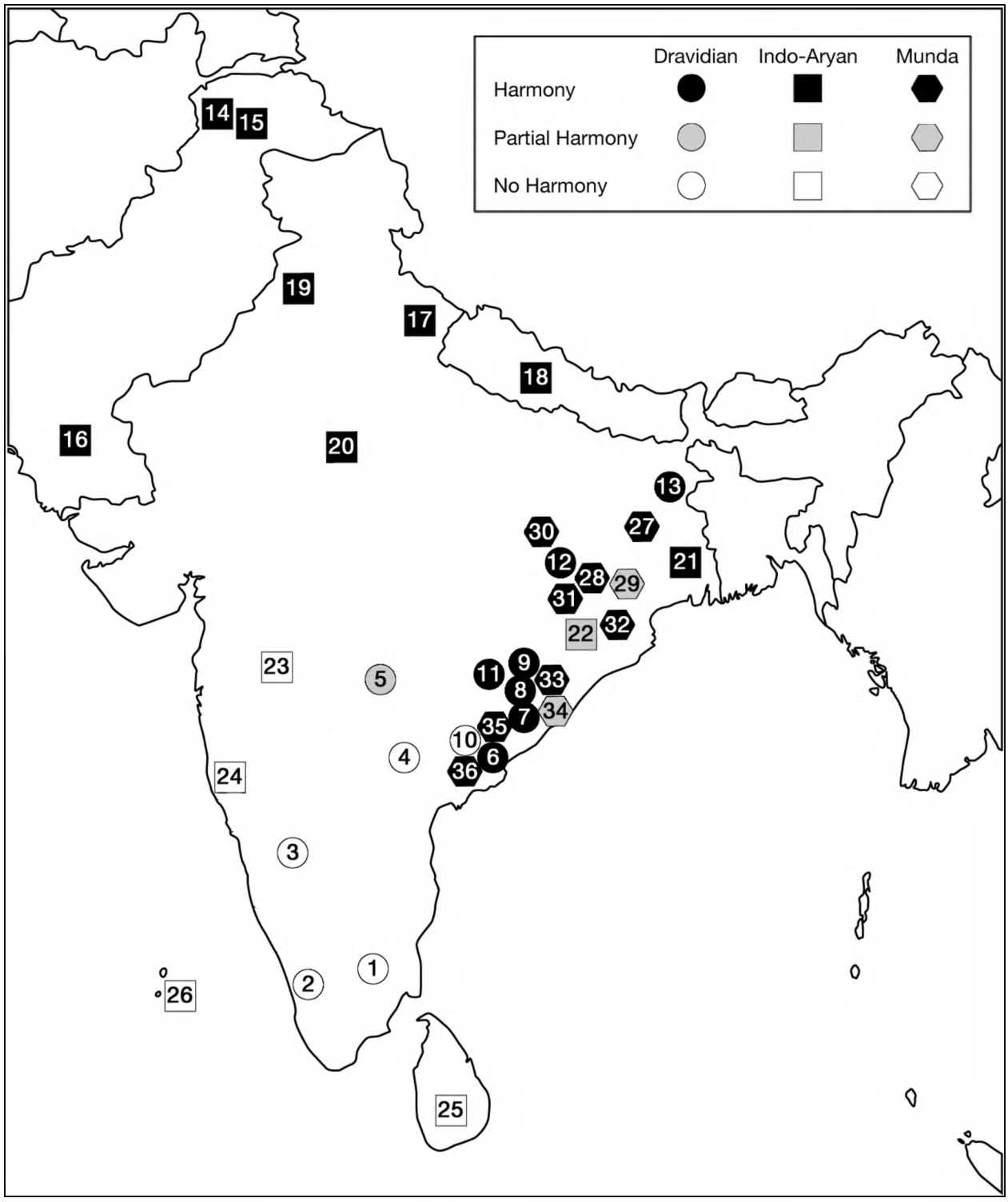

Figure 1 Geographic distribution of languages with and without retroflex consonant harmony in South Asia

Table 17 Index of languages in Figure 1

#	Language	Classification	#	Language	Classification
1	Tamil	Dr, South	19	Panjabi	IA, Central
2	Malayalam	Dr, South	20	Hindi	IA, Central
3	Kannada	Dr, South	21	Bangla	IA, Eastern
4	Telugu	Dr, South-Central	22	Oriya	IA, Eastern
5	Gondi	Dr, South-Central	23	Marathi	IA, Southern
6	Konda	Dr, South-Central	24	Konkani	IA, Southern
7	Kuvi	Dr, South-Central	25	Sinhalese	IA, Sinhalese-Maldivian
8	Pengo	Dr, South-Central	26	Dhivehi	IA, Sinhalese-Maldivian
9	Kui	Dr, South-Central	27	Santali	Munda, North
10	Gadaba	Dr, Central	28	Mundari	Munda, North
11	Parji	Dr, Central	29	Ho	Munda, North
12	Kurux	Dr, North	30	Korwa	Munda, North
13	Malto	Dr, North	31	Kharia	Munda, South
14	Kalasha	IA, Northwestern	32	Juang	Munda, South
15	Indus Kohistani	IA, Northwestern	33	Remo	Munda, South
16	Sindhi	IA, Northwestern	34	Gorum	Munda, South
17	Kumauni	IA, Northern	35	Gta?	Munda, South
18	Nepali	IA, Northern	36	Gutob	Munda, South

As shown in Figure 1, retroflex consonant harmony appears to be an areal feature affecting most languages in the northern half of the sub-continent, outside of the Tibeto-Burman family. Languages with retroflex consonant harmony are concentrated in the north while those that lack harmony are concentrated in the south, regardless of genetic affiliation. Languages with a more gradient form of 'partial' harmony are concentrated in the east. Thus, it appears that retroflex consonant harmony is strongest in the northwest, north and central regions, a little weaker in the east, and absent altogether in the south.

The system of retroflex consonant harmony documented here is remarkably uniform across South Asian languages, despite their genetic diversity and dispersion over a vast geographic area. Before concluding, it is worth summarizing the typological properties of this system.

8. Typological properties

Retroflex consonant harmony in South Asia is remarkably consistent with respect to at least three typological parameters: the domain of harmony, the direction of assimilation and the class of interacting segments. First of all, harmony is always limited to the domain of the lexical root. No language examined here is reported to have morphophonemic alternations as a result of consonant harmony applying across morpheme boundaries.[7] This is not unusual. Root-internal consonant harmony systems are common cross-linguistically (Rose and Walker 2004; Hansson 2010).

Secondly, with respect to direction, the diachronic assimilation that produced harmony was always regressive (T-Ṭ → Ṭ-Ṭ). No examples of progressive assimilation were found (Ṭ-T → Ṭ-

Ṭ). In one sense this is unsurprising given that OIA and Proto-Dravidian both avoided word-initial retroflex consonants. Without Ṭ-T configurations there can be no progressive assimilation. It is important to note, however, that all South Asian languages with retroflex consonant harmony have introduced word-initial retroflex stops independent of consonant harmony. As a result, Ṭ-P, Ṭ-Č and Ṭ-K configurations are all quite common (e.g., Panjabi /ɖabbaː/ 'tin box', /ɖandʒʰ/ 'thirst', /ʈuk-ɳaː/ 'to cut'). In light of this development, the avoidance of Ṭ-T configurations in languages with harmony is significant and suggests the possibility of a bi-directional co-occurrence restriction. Nevertheless, in the absence of historical-comparative evidence of progressive assimilation, all we can say for certain is that South Asian languages with retroflex consonant harmony have systematically avoided introducing Ṭ-T configurations while subjecting inherited T-Ṭ configurations to regressive retroflex assimilation.

Thirdly, in every case, harmony applies primarily (if not exclusively) to co-occurring obstruents with the same manner of articulation. Harmony between co-occurring stops is widespread and systematic but harmony between stops and sonorants is sporadic at best. Long-distance phonological interactions often exhibit similarity effects of this kind; interacting segments are those that are most similar to one another by virtue of sharing many phonological features in common (Rose and Walker 2004; Hansson 2010). The role of similarity is particularly evident in Indo-Aryan languages of the Dardic group, including Indus Kohistani and Kalasha, where consonant harmony applies to co-occurring stops and co-occurring sibilants, but not between stops and sibilants.

Another often cited property of consonant harmony systems is the transparency of intervening segments. Segments that occur between the trigger and target of long-distance assimilation rarely undergo assimilation themselves or block it (Rose and Walker 2004; Hansson 2010). This appears to be true in the present case as well. However, the study looked only at $\#C_1V(N)C_2$ sequences. A more careful examination of harmony over longer domains is required to confirm this observation (e.g., $\#C_1VCVC_2$ roots).

9. Conclusion

In conclusion, we see that many South Asian languages have developed a co-occurrence restriction on coronal obstruents that can be described as retroflex consonant harmony. In these languages, roots containing two non-adjacent coronal stops are primarily limited to those with two dentals (T-T) or two retroflexes (Ṭ-Ṭ), while roots containing a combination of dental and retroflex stops are avoided (*T-Ṭ, *Ṭ-T). This pattern cuts across genetic boundaries affecting most Indo-Aryan, Dravidian and Munda languages in the northern half of the Indian subcontinent, but not those in the south. Thus, it constitutes an areal feature within South Asia.

Retroflex consonant harmony has received little prior attention in the literature on South Asian languages. Previous studies are limited to brief observations concerning specific languages, such as Panjabi (Jain 1934) or Malto (Mahapatra 1979), or groups of closely related languages, such as South-Central Dravidian (Burrow and Bhattacharya 1963). The present study is the first broad and systematic investigation of retroflex consonant harmony in South Asia and (to the best of our knowledge) the first to document it as an areal phenomenon that transcends genetic boundaries in the region. Nevertheless, much work remains to be done. No doubt many cases have gone unnoticed, or at least unreported in the literature. More case studies are required

before we can begin to plot accurate isoglosses on the map and determine their relation to other known isoglosses in the region.

References

Arsenault, Paul. 2012. *Retroflex consonant harmony in South Asia*. Toronto: University of Toronto dissertation.

Arsenault, Paul & Alexei Kochetov. 2011. Retroflex harmony in Kalasha: agreement or spreading? In Suzi Lima, Kevin Mullin & Brian Smith (eds.), *Proceedings of the North East Linguistic Society* 39. 55–66. Amherst: GLSA.

Bhaduri, Manindra Bhusan. 1983 [1931]. *A Mundari-English dictionary*. New Delhi: Northern Book Centre.

Bhaskararao, Peri. 1980. *Konekor Gadaba: A Dravidian language*. Pune: Deccan College Post-Graduate & Research Institute.

Bhat, D. N. S. 1973. Retroflexion: An areal feature. *Working Papers on Language Universals* 13. 27–67.

Biswas, Sailendra. 2000. *Samsad Bengali-English dictionary*, 3rd edn. Calcutta: Sahitya Samsad. http://dsal.uchicago.edu/dictionaries/biswas-bengali/.

Bodding, P. O. 1929–1936. *A Santali dictionary*, 5 vols. Oslo: I Kommisjon Hos Jacob Dybwab.

Burrow, T. & S. Bhattacharya. 1953. *The Parji language: A Dravidian language of Bastar*. Hertford: Stephen Austin and Sons.

Burrow, T. & S. Bhattacharya. 1961. Some notes on the Kui dialect as spoken by the Kuṭṭia Kandhs of North-East Koraput. *Indo-Iranian Journal* 5. 118–35.

Burrow, T. & S. Bhattacharya. 1963. Notes on Kuvi with a short vocabulary. *Indo-Iranian Journal* 6. 231–289.

Burrow, T. & S. Bhattacharya. 1970. *The Pengo language: Grammar, texts, and vocabulary*. Oxford: Oxford University Press.

Burrow, T. & M. B. Emeneau. 1984. *A Dravidian etymological dictionary*, 2nd edn. Oxford: Clarendon Press. http://dsal.uchicago.edu/dictionaries/burrow/.

Coetzee, Andries W. & Joe Pater. 2008. Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language and Linguistic Theory* 26. 289–337.

Deeney, J. 1978. *Ho-English dictionary*. Chaibasa: Xavier Ho Publications.

Donegan, Patricia J. & David Stampe. 2004. *Munda lexical archive*. http://www.ling.hawaii.edu/austroasiatic/ (accessed 24 August, 2010).

Emeneau, M. B. 1956. India as a linguistic area. *Language* 32(1). 3–16.

Fabricius, Johann Philipp. 1972. *J. P. Fabricius's Tamil and English dictionary*, 4th edn. Tranquebar: Evangelical Lutheran Mission Pub. House. http://dsal.uchicago.edu/dictionaries/fabricius/.

Frisch, Stefan A., Janet B. Pierrehumbert & Michael B. Broe. 2004. Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* 22(1). 179–228.

Gallagher, Gillian & Jessica Coon. 2009. Distinguishing total and partial identity: Evidence from Chol. *Natural Language and Linguistic Theory* 27(3). 545–582.

Goswami, Krishan Kumar. 2000. *Panjabi-English English-Panjabi dictionary.* New Delhi: UBS Publishers' Distributors.
Grignard, A. 1986 [1924]. *An Oraon-English dictionary.* New Delhi: Unity Book Service.
Gwynn, John P. L. 1991. *A Telugu-English dictionary.* Delhi & New York: Oxford University Press. http://dsal.uchicago.edu/dictionaries/gwynn/.
Hansson, Gunnar Ólafur. 2010. *Consonant harmony: Long-distance interaction in phonology.* Berkeley: University of California Press.
Israel, M. 1979. *A grammar of the Kuvi language (with texts and vocabulary).* Trivandrum: Dravidian Linguistics Association.
Jain, B. D. 1934. *A phonology of Panjābī as spoken about Ludhiana and a Ludhiānī phonetic reader*. Lahore: The University of the Panjab.
Kawahara, Shigeto, Hajime Ono & Kiyoshi Sudo. 2006. Consonant cooccurrence restrictions in Yamoto Japanese. *Japanese and Korean Linguistics* 14. 27–38.
Krishnamurti, Bhadriraju. 1969. *Konda or Kubi: A Dravidian language.* Hyderabad: Tribal Cultural Research & Training Institute, Government of Andhra Pradesh.
Krishnamurti, Bhadriraju. 2003. *The Dravidian languages.* Cambridge: Cambridge University Press.
Lewis, M. Paul (ed.). 2009. *Ethnologue: Languages of the world*, 16th edn. Dallas: SIL International.
Maffei, Angelus Francis Xavier. 1883. *A Konkani-English dictionary.* Mangalore: Basel Mission Press.
Mahapatra, B. P. 1979. *Malto – An ethnosemantic study*. Mysore: Central Institute of Indian Languages.
Mahapatra, B. P. 1987. *Malto-Hindi-English dictionary.* Mysore: Central Institute of Indian Languages.
Malik, Amar Nath. 1995. *The phonology and morphology of Panjabi.* New Delhi: Munshiram Manoharlal Publishers.
Masica, Colin P. 1991. *The Indo-Aryan languages.* Cambridge: Cambridge University Press.
McGregor, R. S. 1993. *The Oxford Hindi-English dictionary.* Oxford: Oxford University Press.
Molesworth, J. T. 1857. *A dictionary, Marathi and English*, 2nd edn. Bombay: Printed for government at the Bombay Education Society's press. http://dsal.uchicago.edu/dictionaries/molesworth/.
Penny, Mark, Joanna Penny & Pendur Durnath Rao. 2005. *Gondi dictionary.* Utnoor, AP: Integrated Tribal Development Agency.
Peterson, John. 2009. A Kharia-English lexicon. *Himalayan Linguistics Archive* 5. i–xv, 1–212. http://www.linguistics.ucsb.edu/HimalayanLinguistics/grammars/2009/HLA05_Peterson.pdf. (accessed 29 March, 2011).
Pfeiffer, Martin. 1972. *Elements of Kurux historical phonology.* Leiden: E. J. Brill.
Pozdniakov, Konstantin & Guillaume Segerer. 2007. Similar place avoidance: A statistical universal. *Linguistic Typology* 11. 307–348.
Ramanujan, A. K. & Colin Masica. 1969. Toward a phonological typology of the Indian linguistic area. In T. A. Sebeok (ed.), *Current trends in linguistics, volume 5: Linguistics in South Asia*, 543–577. Paris: Mouton.

Reynolds, Christopher. 2003. *A Maldivian dictionary.* London & New York: RoutledgeCurzon.
Rose, Sharon, & Rachel Walker. 2004. A typology of consonant agreement as correspondence. *Language* 80(3). 475–531.
Schwarzschild, L. A. 1973. Initial retroflex consonants in Middle Indo-Aryan. *Journal of the American Oriental Society* 93(4). 482–487.
Subrahmanyam, P. S. 1983. *Dravidian comparative phonology.* Annamalainagar: Annamalai University.
Turner, R. L. 1931. *A comparative and etymological dictionary of the Nepali language.* London: K. Paul, Trench, Trubner. http://dsal1.uchicago.edu/dictionaries/turner/.
Turner, R. L. 1962–1966. *A comparative dictionary of Indo-Aryan languages.* London: Oxford University Press. http://dsal.uchicago.edu/dictionaries/soas/.
Turner, R. L. 1969. *A comparative dictionary of Indo-Aryan languages. Indexes.* London: Oxford University Press.
Winfield, W. W. 1928. *A grammar of the Kui language.* Calcutta: Asiatic Society of Bengal.
Zoller, Claus Peter. 2005. *A grammar and dictionary of Indus Kohistani, volume 1: Dictionary.* Berlin & New York: Mouton de Gruyter.
Zvelebil, Kamil V. 1970. *Comparative Dravidian phonology.* The Hague: Mouton.

---

[1] /ṁ/ is a common romanization of orthographic *anuswara* in Indic scripts. In MIA it may have represented nasalization of the preceding vowel or a homorganic nasal consonant.

[2] The only zone not represented in the study is the East-Central zone, which consists of five languages at most (Lewis 2009). Insufficient data was available for these languages.

[3] In Pozdniakov and Sergerer (2007) the thresholds are ±15% and ±30% instead of ±25% and ±50%, respectively. However, the lower thresholds fail to capture the distinction between a language with near-categorical avoidance of a given C_1-C_2 pair (e.g., −95%) and one with a more gradient avoidance (e.g., −30%). By increasing the thresholds, each category (single vs. double plus/minus) is reserved for pairs that show a stronger discrepancy between observed and expected counts.

[4] Turner (1962–1966) and Goswami (2000) have /ʈɖʰaː/ but Jain (1934) has /ʈɖɖʰaː/.

[5] This complication does not entail that the statistical method is incapable of distinguishing between languages with and without retroflex consonant harmony. The two types of language are distinguished not so much by their O/E values for Ṭ-Ṭ configurations, but by their O/E values for T-Ṭ configurations. In languages that avoid word-initial retroflexes the O/E ratio for T-Ṭ configurations is always very close to 1.0, indicating that disharmonic configurations are not avoided in favour of harmonic ones (cf. Oriya in Table 5). In contrast to this, languages with retroflex consonant harmony always have very low O/E ratios for T-Ṭ configurations because they have been subjected to harmony.

[6] Malto /ʈoʈr-/ ‘to grope’ might be cognate with Tamil /toʈu/ ‘to touch’ and other items listed under DEDR 3480. Cf. Parji /ʈoɖ-, ʈoʈʈ-/ ‘to touch’ in Table 13.

[7] Long-distance retroflex assimilation may be responsible for alternations in Sanskrit (via the n-retroflexion rule), Burushaski (Isolate) and Sherpa (Tibeto-Burman). However, none of these

patterns appear to be connected with the system of consonant harmony described here. See Arsenault (2012) for discussion.