Retroflex Consonant Harmony in South Asia

by

Paul Edmond Arsenault

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy

Department of Linguistics
University of Toronto

Retroflex Consonant Harmony in South Asia

Paul Edmond Arsenault

Doctor of Philosophy

Department of Linguistics
University of Toronto

2012

Abstract

This dissertation explores the nature and extent of retroflex consonant harmony in South Asia. Using statistics calculated over lexical databases from a broad sample of languages, the study demonstrates that retroflex consonant harmony is an areal trait affecting most languages in the northern half of the South Asian subcontinent, including languages from at least three of the four major families in the region: Dravidian, Indo-Aryan and Munda (but not Tibeto-Burman). Dravidian and Indo-Aryan languages in the southern half of the subcontinent do not exhibit retroflex consonant harmony.

In South Asia, retroflex consonant harmony is manifested primarily as a static co-occurrence restriction on coronal consonants in roots/words. Historical-comparative evidence reveals that this pattern is the result of retroflex assimilation that is non-local, regressive and conditioned by the similarity of interacting segments. These typological properties stand in contrast to those of other retroflex assimilation patterns, which are local, primarily progressive, and not conditioned by similarity. This is argued to support the hypothesis that local feature spreading and long-distance feature agreement constitute two independent mechanisms of assimilation, each with its own set of typological properties, and that retroflex consonant harmony is the product of agreement, not spreading. Building on this hypothesis, the study offers a formal account of retroflex consonant harmony within the Agreement by Correspondence (ABC) model of Rose & Walker (2004) and Hansson (2001; 2010).

Two Indo-Aryan languages, Kalasha and Indus Kohistani, figure prominently throughout the dissertation. These languages exhibit similarity effects that have not been clearly observed in other retroflex consonant harmony systems; retroflexion is contrastive in both non-sibilant (i.e., plosive) and sibilant obstruents (i.e., affricates and fricatives), but harmony applies only within each manner class, not between them. At the same time, harmony is not sensitive to laryngeal features. Theoretical implications of these and other similarity effects are discussed.

Acknowledgments

Many people have assisted me in one way or another throughout the process of researching and writing this dissertation. The members of my thesis committee deserve special mention. Keren Rice is everything one could want in a supervisor. She was always available, always a good listener, always a source of excellent advice (linguistic and otherwise) and always had my best interest at heart. It has been a joy and privilege to study under her guidance. Alexei Kochetov has contributed more than anyone else to the substance of the dissertation. It was he that first put me onto the topic of consonant harmony. Several of the case studies presented in this dissertation were conducted in collaboration with him, including Kalasha (§3.3.2), Panjabi (§3.2.3) and Burushaski (§3.3.5). Elan Dresher always had insightful and constructive feedback. There is no doubt that the dissertation is stronger for his input.

I am also very grateful to Gunnar Hansson (University of British Columbia), Yoonjung Kang (University of Toronto Scarborough) and Peter Avery (York University) for agreeing to read what is, admittedly, a long dissertation, and for serving on the examination committee. Their questions and comments in the final oral examination were both stimulating and constructive and will no doubt set the course for my research in years to come.

For me, one of the most rewarding aspects of this research project was the opportunity it afforded for collaboration with linguists from around the globe. Several people graciously contributed electronic lexical databases from their own fieldwork on lesser-known and under-documented languages of South Asia. The main contributors, and the languages for which they contributed data, are as follows (in no particular order): Ron Trail and Greg Cooper (Kalasha); Joan Baart (Kalami); Henrik Liljegren (Palula); the late Carla Radloff (Shina); Irena van Riezen

(Kumauni); Binzy Joseph George and Christina Joseph (Korwa); Laiju Ek (Bagheli). Others shared research in the form of papers or took the time to discuss their area of specialization with me. In addition to some of the names just mentioned, these include (again, in no particular order): Jan Heegård and Ida Mørch (Kalasha); Pierpaolo Di Carlo (Kalasha); Claus Peter Zoller (Indus Kohistani); Rajesh Khatiwada (Nepali); Christopher Wilde (Humla Bhotia); Kevin Kopp (Dolpo Tibetan); and Stella Sandahl (Sanskrit).

Parts of this research project were presented at various conference venues, where I received valuable feedback from fellow participants. My thanks to the participants at the 39th meeting of the North East Linguistic Society (NELS 39) at Cornell University in 2008; the Linguistic Society of America (LSA) in San Francisco in 2009; the MOT Phonology Workshop at the University of Toronto in 2009 and 2012; the CRC-Sponsored Phonetics/Phonology Workshop at the University of Toronto in 2010 and 2011; and the Canadian Linguistic Association (CLA) at Wilfrid Laurier University in 2012. In particular, I benefited greatly from discussions, mostly concerning theoretical approaches to consonant harmony, with Rachel Walker, Sara Mackenzie, Daniel Currie Hall and Peter Jurgec, among others.

Thanks to my immediate and extended family who supported me in so many practical ways. Special thanks to John Bateman for many early morning drives to the train station and to Victor Ettel and Sue Perry-Ettel for giving me a place to rest my head in Toronto. Last, but not least, I thank my wife, Kyla, and our children, TJ and Hannah, for putting up with my long absences from home and with my absent mindedness when I was at home. More than anyone or anything else, they kept me grounded when the going got tough by reminding me of all that I love in the world outside of linguistics.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

AA	Austro-Asiatic
ABC	Agreement by Correspondence
Bsk	Burushaski
CDIAL	Comparative Dictionary of Indo-Aryan Languages (Turner, 1962–1966)
DEDR	Dravidian Etymological Dictionary (Revised) (Burrow & Emeneau, 1984)
Dr	Dravidian
DT	Dispersion Theory
Go	Gondi
IA	Indo-Aryan
IE	Indo-European
IIr	Indo-Iranian
IK	Indus Kohistani
IPA	International Phonetic Alphabet
LDCA	Long-Distance Consonant Assimilation
Ma	Malayalam
MIA	Middle Indo-Aryan
MSC	Morpheme Structure Constraint
NIA	New Indo-Aryan
OIA	Old Indo-Aryan
OT	Optimality Theory
PIE	Proto-Indo-European

Pkt	Prakrit
SDA	Successive Division Algorithm
Sh	Shina
Skt	Sanskrit
Ta	Tamil
TB	Tibeto-Burman
VOT	Voice Onset Time

A note on transcription conventions

This thesis draws on data from a wide range of sources employing a variety of different transcription conventions. For consistency, all transcriptions have been modified to bring them in line with the conventions of the International Phonetic Alphabet (IPA). The most common convention in the source materials is the traditional South Asian transcription system, which developed as a Romanized transliteration for South Asian scripts (Masica, 1991, p. xv). The main correspondences between this traditional system and the IPA system are summarized here.

Within the South Asian tradition, retroflexion is typically represented by means of an underdot, as in *ṭ*, *ḍ*, *ṣ*, *ẓ*, *ṇ*, *ṛ*, *ḷ*. These have been modified to IPA ‘right-tail’ characters: *ʈ*, *ɖ*, *ʂ*, *ʐ*, *ɳ*, *ɽ*, *ɭ*, respectively. The traditional underdot is sometimes extended to retroflex affricates, giving transcriptions such as *c̣*, *j̣*, or *č̣*, *ǰ̣*. These have been modified to IPA *ʈʂ* and *ɖʐ*. Retroflex approximants are transcribed variously as *r̤*, *ṛ*, *ḻ*, *ẓ* or *zh* (among other things) in transcriptions of Dravidian languages, and as *ỵ* in transcriptions of Burushaski. All of these have been modified to IPA *ɻ*. Trail & Cooper (1999) transcribe the retroflex vowels of Kalasha as *ạ*, *ị*, *ụ*, *ẹ*, *ọ*, etc. These have been modified to *a˞*, *i˞*, *u˞*, *e˞*, *o˞*, etc., using the IPA diacritic for rhoticity.

Some Dravidian languages distinguish apico-alveolar stops and nasals from their lamino-dental counterparts. In traditional transcriptions, the dentals are typically unmarked (*t*, *d*, *n*) while the alveolars are distinguished by means of an underbar (*ṯ*, *ḏ*, *ṉ*). As a general rule, I have retained the practice of leaving dentals unmarked, although in a few places I have used IPA *t̪*, *d̪*, *n̪* for clarity. Apico-alveolars have been modified to *t̺*, *d̺*, *n̺*, using the IPA diacritic for apicality. In the Dravidian literature, the voiceless apico-alveolar stop is often represented phonemically as */ṟ/* (instead of */ṯ/*) because it is typically a voiced trill intervocalically. Nevertheless, it generally remains a voiceless stop in gemination and a voiced stop after nasals. Thus, for transparency, I have modified *ṟṟ* and *ṉṟ* to IPA /*t̪t̪*/ and /*n̪t̪*/ in phonemic representations and to [*t̪t̪*] and [*n̪d̪*] in phonetic representations.

Most South Asian languages have a series of so-called ‘palatal’ stops that are realized as laminal post-alveolar affricates. These are typically transcribed as *c*, *č* or *ć*, in the case of voiceless phonemes, and *j* or *ǰ*, in the case of their voiced counterparts. I have modified these to IPA palato-alveolar affricates, *ʧ* and *ʤ*, throughout. Similarly, ‘palatal’ post-alveolar fricatives, which are typically transcribed as *š* or *ś* and *ž* in the literature, have been modified to IPA *ʃ* and *ʒ*. The IPA alveolo-palatals, *ʨ*, *ʥ*, *ɕ* and *ʑ*, are also possible equivalents for palatal affricates and fricatives in many South Asian languages. For the sake of consistency I have only used the palato-alveolars, *ʧ*, *ʤ*, *ʃ* and *ʒ*. Within the literature, *ċ* and *j̇* are sometimes used for dental or alveolar affricates. These have been modified to IPA *ʦ* and *ʣ*. The use of *j* throughout this dissertation corresponds to a palatal approximant (typically *y* in the source materials), not a voiced palatal stop/affricate, as in the South Asian tradition.

The use of IPA *n* (or *n̪*), *n̪* and *ɳ* for traditional *n*, *ṉ* and *ṇ* has already been noted. In addition, traditional transcriptions employ *ñ* for palatal nasals and, occasionally, *ṅ* for velar nasals. These have been replaced with IPA *ɲ* and *ŋ*, respectively. In the South Asian transliteration system, the symbol *ṁ* corresponds to orthographic *chandrabindu* (ँ), which typically denotes nasalization of the preceding vowel. Thus, sequences such as *aṁ*, *uṁ*, etc., have been modified to IPA *ã*, *ũ*, etc.

In the South Asian tradition, *ph*, *bh*, *th*, *dh*, *kh*, *gh*, etc., represent aspirated consonants, not sequences of C + *h*. Following IPA conventions, these have been transcribed as p^h, b^h, t^h, d^h, k^h, g^h, etc. Strictly speaking, the voiced aspirated stops are realized with breathy voice. As such, they could be represented with superscript IPA *ɦ*, as in $b^{ɦ}$, $d^{ɦ}$, $g^{ɦ}$, etc. For simplicity, I have retained the use of superscript *h*.

A macron is typically used to represent vowel length in the South Asian literature. Thus, *ā*, *ī*, *ū*, *ē* and *ō* have been modified to IPA *aː*, *iː*, *uː*, *eː* and *oː*. In addition, the vocalic (i.e., syllabic) liquid of Sanskrit, which is traditionally represented as *r̥* (or occasionally *ṛ*), has been converted to IPA *r̩* in the present study, using the IPA diacritic for syllabic segments.

Chapter 1
Introduction

Consonant harmony is a relatively understudied subject, both in the context of South Asia and in the field of phonology in general. Until recently, it was regarded as a phenomenon primarily affecting coronal consonants, most notably coronal sibilants, in a relatively small number of languages (Gafos, 1999). However, recent cross-linguistic surveys by Hansson (2001; 2010) and Rose & Walker (2004) have revealed that consonant harmony is much more common than previously suspected and affects a wider range of segments and features than those of the coronal class. A new and greater awareness of consonant harmony is stimulating research into this previously neglected area, and new cases and details are coming to light (Sibanda, 2004; Martin, 2005; Kochetov, 2007; Brown, 2008; 2010; Arsenault, 2009a; Arsenault & Kochetov, 2011; Gallagher, 2010; 2012). The present study contributes to this growing body of knowledge by exploring the nature and extent of retroflex consonant harmony in South Asian languages.

The study of retroflex consonant harmony in South Asia has much to offer, both empirically and theoretically. To date, the most comprehensive cross-linguistic survey of consonant harmony systems is that of Hansson (2010), which includes approximately 175 separate cases, drawn from more than 130 languages, representing a broad range of geographic regions and language families. Only five of these cases are from South Asia.[1] From this, it

[1] An earlier version of Hansson's survey (Hansson, 2001) included only four cases from South Asia: retroflex and dorsal place harmonies in Malto (Dravidian), 'palatal' place harmony in Pengo (Dravidian) and laryngeal harmony in Gojri (Indo-Aryan). Hansson (2010) includes a fifth case: retroflex consonant harmony in Kalasha (Indo-Aryan), based on preliminary results from the present study reported in Arsenault & Kochetov (2009; 2011).

might be possible to infer that consonant harmony is indeed a rare phenomenon in South Asia, if not elsewhere, and that the few cases cited by Hansson are exceptional. This is not the case. A major empirical finding of the present study is that retroflex consonant harmony is a widespread areal trait affecting most languages in the northern half of the South Asian subcontinent, including languages from at least three of the four major South Asian families: Dravidian, Indo-Aryan and Munda. Moreover, some of these languages exhibit striking typological properties that have not been clearly observed in other retroflex consonant harmony systems. These empirical properties have implications for phonological theories, which must provide an account of the attested sound patterns. Thus, the study of retroflex consonant harmony in South Asia has much to offer on every level.

The goals of this dissertation are both empirical and theoretical. Empirically, the study seeks to address questions such as the following: How common is retroflex consonant harmony in South Asia? Which South Asian languages exhibit retroflex consonant harmony and which do not? What typological properties does retroflex consonant harmony exhibit in South Asia, and how do those properties compare with the typological properties of other consonant harmony systems cross-linguistically? Of the many theoretical issues that arise from the study of consonant harmony, only two are discussed in any depth. First, what kind of assimilatory mechanism is responsible for consonant harmony, and does the same mechanism also drive

---

Hansson (2001; 2010) also discusses n-retroflexion in Sanskrit (Indo-Aryan) but argues that it is not a case of consonant harmony, as defined in much contemporary work on the subject, because it displays typological properties that are not consistent with those of other consonant harmony systems. See §3.2.1 for further discussion.

other assimilation patterns, or is it something unique to consonant harmony? Second, to what extent does the similarity of interacting segments determine their participation in consonant harmony, and how is the similarity of interacting segments evaluated?

Before proceeding with these questions, it is necessary to define some key terms, concepts, and research methods that are foundational to the rest of the study. This introductory chapter lays the groundwork for subsequent chapters by addressing such questions as: What are the languages of South Asia? What is retroflexion? And what is consonant harmony? Section §1.1 delimits the geographic region of South Asia, and provides an overview of the language families in the region, their genetic sub-classification, and their geographic distribution. Section §1.2 summarizes the articulatory properties and acoustic/perceptual cues associated with retroflexion. The properties of retroflex segments are discussed in relation to those of other segments within the coronal class, to which they belong. Consonant harmony is defined in §1.3, and the typological parameters relevant to the study of consonant harmony systems are discussed. Some statistical methods used to evaluate consonant harmony systems are briefly explained in §1.4. Finally, §1.5 provides a brief overview of the dissertation.

1.1 Languages of South Asia

This dissertation is concerned primarily with the languages of South Asia. South Asia is a geographic region that encompasses the countries of India, Pakistan, Nepal, Bhutan, Bangladesh, Sri Lanka, and Maldives. Four major language families are represented in the region: (i) Dravidian; (ii) Indo-Iranian (a sub-family of Indo-European), which includes Indo-Aryan and Iranian languages; (iii) Austro-Asiatic, which includes Munda and Mon-Khmer languages; and (iv) Tibeto-Burman (a sub-family of Sino-Tibetan). The Andamanese family, a few Tai-Kadai languages and a handful of isolates (e.g., Burushaski, Kusunda and Nihali) also fall within the borders of South Asia. All of these language families have been in contact on the South Asian subcontinent since prehistoric times. As a result, South Asia now constitutes a ‘linguistic area’ or ‘Sprachbund’; a geographic area in which languages of different genetic stock have come to resemble one another through a history of contact and convergence (Emeneau, 1956; Masica, 1976).

The vast majority of Indo-Iranian languages in South Asia are Indo-Aryan, as opposed to Iranian. As a result, any generalizations made about Indo-Iranian languages within South Asia are primarily generalizations concerning the Indo-Aryan branch. Similarly, the majority of Austro-Asiatic languages in South Asia are Munda, as opposed to Mon-Khmer. As a result, generalizations about Austro-Asiatic languages within South Asia are essentially generalizations concerning the Munda branch. For these reasons, it is common to find Indo-Aryan and Munda listed in *lieu* of Indo-Iranian and Austro-Asiatic, respectively, in the South Asian literature. As a general rule, this practice is adopted throughout the present study.

Altogether, approximately 539 languages are spoken in South Asia. Language and population figures for each South Asian family are summarized in Table 1. The geographic distribution of the families is shown on the map in Figure 1.

Table 1 Languages of South Asia[2]

	Number of Languages (sub-totals)		Number of Languages (totals)	Number of Speakers (in millions)
Indo-Iranian	*Indo-Aryan*	184	191	1099
	Iranian	7		
Dravidian	–		84	222
Austro-Asiatic	*Munda*	22	32	12
	Mon-Khmer	10		
Tibeto-Burman	–		218	16
Other	*Andamanese*	4	14	3
	Tai-Kadai	4		
	Isolate	3		
	Unclassified	3		
Grand Totals			539	1352

[2] The figures in Table 1 are calculated on the basis of information in the 16th edition of the *Ethnologue* (Lewis, 2009). With the exception of Dravidian, Munda and Andamanese, the language figures do not represent the total number of languages in each family, but rather the total number of languages from each family (or sub-family) spoken in the South Asia region. These figures do not include three European languages (English, Portuguese and French), six pidgins/creoles, four sign languages and thirteen extinct languages (most of them Andamanese or Tibeto-Burman), all of which are also listed for the countries of South Asia. If these are included, then the total number of languages for the region is 565. For a more conservative estimate of figures see Asher (2008, p. 33).

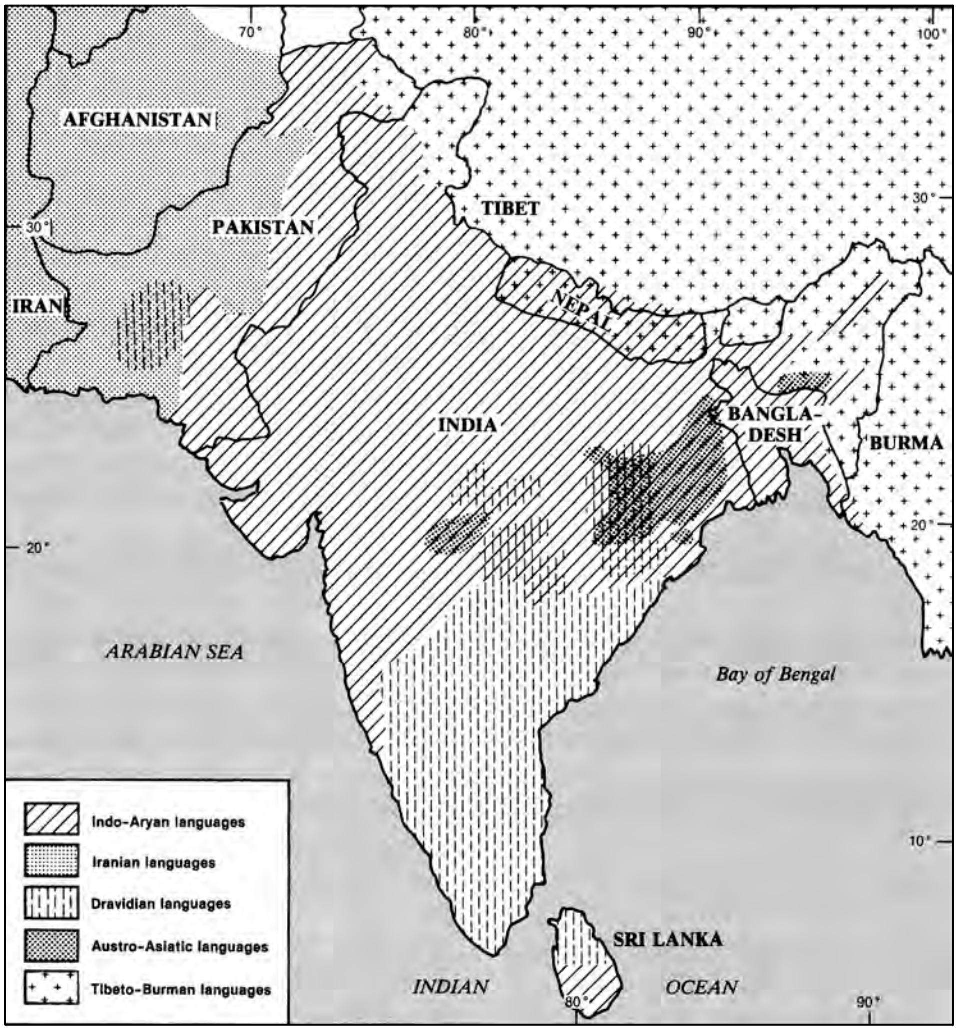


Figure 1 Distribution of language families in South Asia (Masica, 1992)

The Indo-Aryan family is by far the largest in terms of population and geographic spread. It dominates the northern half of the Indian subcontinent and accounts for over 80% of the South Asian population. The Dravidian family dominates the southern half of the subcontinent and is the second largest family population-wise. The Munda family consists of a small number of minority languages in eastern India. Tibeto-Burman is the largest family in terms of distinct language varieties but is among the smallest in terms of population. It is restricted to the Himalayan region on the northern and northeastern peripheries of the subcontinent but also extends to adjacent areas well beyond the confines of South Asia.

At one time, the Munda languages were probably spoken over a larger geographic area than that which they currently occupy. The speakers of those languages may still represent autochthonous populations in the areas where they are currently spoken (Anderson, 2008b). The Indo-Aryan languages are a late addition to the area. Indo-Aryan populations migrated into South Asia from the northwest sometime around 1500 BCE, displacing or absorbing other people groups in their path (Masica, 1991, p. 37). The story of Dravidian prehistory is uncertain. Whatever their ultimate origin, it is generally accepted that Dravidian speakers already occupied a large part of the subcontinent before the Indo-Aryans arrived, and were among the people that the Indo-Aryans encountered upon their arrival (Krishnamurti, 2003).[3]

In subsequent chapters, generalizations are sometimes stated in relation to specific subgroups within each language family. For ease of reference, the subclassification schemes assumed for the families are presented below in §1.1.1–1.1.4. The sub-classification schemes adopted here are those of the *Ethnologue* (Lewis, 2009, *16th edition*). This is purely a matter of convenience; nothing critical hinges on the choice of these particular schemes over any others. As a general rule, only major subgroups are presented; details concerning minor subgroups are omitted unless they are relevant to the discussion in the following chapters. With few exceptions, the list of languages under a given subgroup is not exhaustive. Languages are listed if they are major representatives of the group or if they figure in later chapters.

---

[3] For an alternative view, see Witzel (1999), who argues that Dravidian speakers immigrated into South Asia through the Sindh region of modern day Pakistan, from somewhere in the vicinity of Iran, around the same time that the Indo-Aryans were immigrating through the northwest. Witzel argues that contact between the two groups occurred at a later date than normally assumed, and that Dravidians were not among the earliest people groups encountered by the Indo-Aryans in South Asia. See also Krishnamurti (2003, pp. 37–38) for a critique of Witzel.

1.1.1 Dravidian languages

The sub-classification of Dravidian languages is relatively uncontroversial. All accounts recognize four major subgroups, typically labeled according to their relative geographic position: South, South-Central, Central and North, as shown in Figure 2.[4]

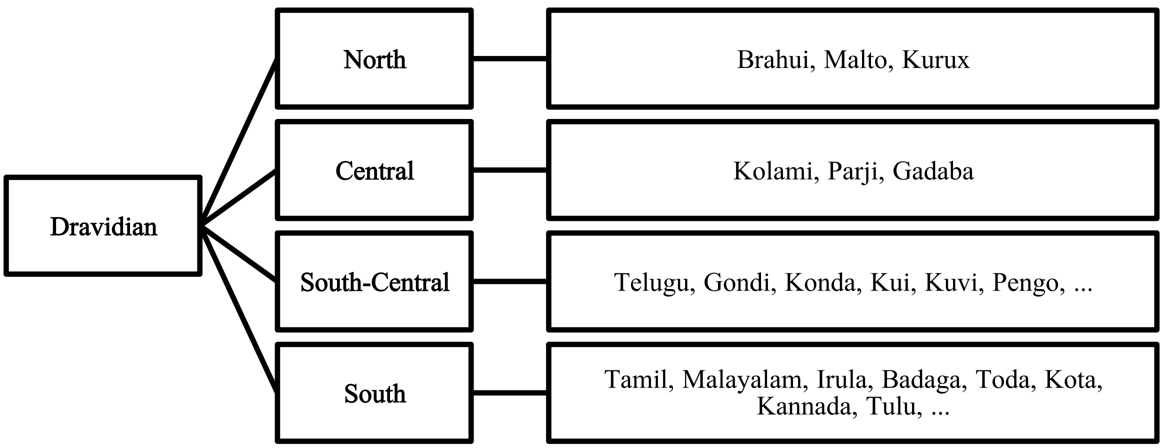


Figure 2 Sub-classification of Dravidian languages

South Dravidian is by far the largest subgroup, accounting for more than half of all Dravidian languages. This subgroup includes three out of four major literary Dravidian languages: Tamil, Malayalam and Kannada. Of these, Tamil has the longest literary tradition, with written records dating back more than two thousand years (Lehmann, 1998). Many languages of this group are also phonologically conservative, preserving elements of Proto-Dravidian that have been lost elsewhere.

---

[4] In some accounts, South Dravidian and South-Central Dravidian are labelled “South Dravidian I” and “South Dravidian II”, respectively (e.g., Krishnamurti, 2003).

The South-Central group is the second largest subgroup after South Dravidian. It includes Telugu, which is the only major literary language outside of the South Dravidian group. The Central and North Dravidian groups are both very small by comparison, consisting of just three to five languages each, depending on how languages and dialects are defined.

1.1.2 Indo-Aryan languages

The Indo-Aryan family has a long literary tradition that preserves specimens of language spanning a period of approximately 3500 years.[5] As a result, Indo-Aryan languages are classified not only according to genetic or geographic subgroups, but also according to one of three broad diachronic periods. These are summarized in (1) along with representative examples of each period.

(1) Diachronic classification of Indo-Aryan languages (Masica, 1991, pp. 50–55)

a. Old Indo-Aryan (OIA): 1500 BCE – 600 BCE
   - Vedic Sanskrit
   - Classical Sanskrit

b. Middle Indo-Aryan (MIA): 600 BCE – 1000 CE
   - Pāli
   - “Prakrits”

---

5 The earliest Indo-Aryan inscriptions are Aśokan Prakrit inscriptions, which date from the third century BCE. However, Vedic Sanskrit texts, which survive only in the form of later manuscripts, were probably composed orally sometime around 1500 BCE, and are believed to preserve the language of that period (Deshpande, 1992).

c. New Indo-Aryan (NIA): 1000 CE – present

- Hindi/Urdu
- Bangla (Bengali)
- Panjabi, etc.

Sanskrit is representative of Old Indo-Aryan. Vedic Sanskrit is the most archaic form of the language. It is preserved in the *Rig Veda*, a collection of hymns, prayers and incantations, dating back to as early as 1500 BCE. Classical Sanskrit represents the later half of the OIA period, as reflected in works such as Pāṇini's grammar, the Aṣṭādhyāyī (c. 500 BCE).

The Middle Indo-Aryan period is represented by Pāli and by various "Prakrits". Pāli is the language of the Hinayana Buddhist canon and other related literature. Prakrit is a cover term for a collection of vernacular dialects found in inscriptions and various texts dating from the MIA period. Prakrit dialects include Aśokan Prakrit, Niya Prakrit, Ardhamāgadhī, Māgadhī, Śaurasenī, Mahārāṣṭrī and others (Masica, 1991, pp. 51–53; cf. Turner, 1969, p. vii).

The New Indo-Aryan period is represented by the broad spectrum of Indo-Aryan languages spoken in present day South Asia. Identifying genetic subgroups within NIA is complicated by the fact that "the entire Indo-Aryan realm (except for Sinhalese) constitutes one enormous dialectal continuum" in which "the speech of each village differs slightly from the next… all the way from Assam to Afghanistan" (Masica, 1991, p. 25, cf. p. 446). Differences between non-contiguous speech varieties on the continuum are significant, but delimiting clear boundaries between one language and the next, or one subgroup and the next, is very difficult. As a result, most accounts classify the various Indo-Aryan languages according to broad geographic zones, though the precise number of zones and their boundaries vary somewhat from one account to another. The major components of the sub-classification scheme assumed here are presented in Figure 3.[6]

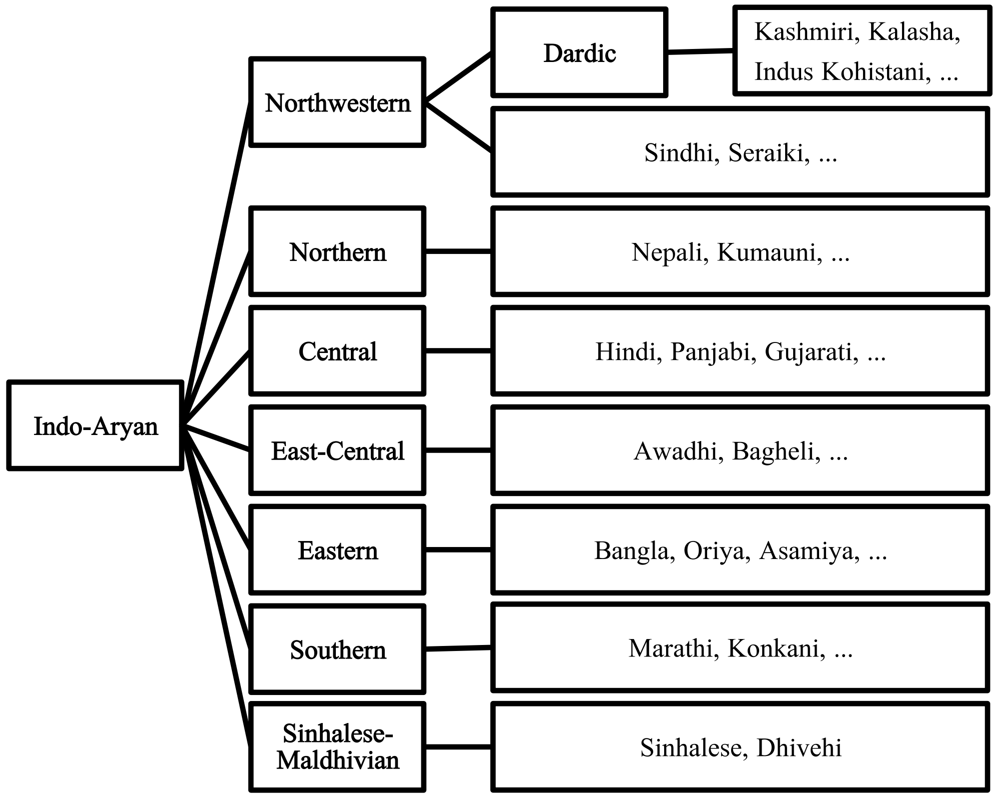


Figure 3 Sub-classification of Indo-Aryan languages

All of the major NIA geographic zones include one or more major literary languages; many of them recognized as official state languages in their respective areas. With the possible exception of the Sinhalese-Maldivian group, most zones also include a large number of minor and/or non-literary languages. The Dardic subgroup of the Northwestern zone deserves special mention. The best-known representative of this group is Kashmiri, a major literary language of

---

6 For a useful summary and review of various alternative NIA subclassification schemes, see Masica (1991, Appendix II, pp. 446–463).

northwest India. However, the Dardic group also includes a number of lesser-known minority languages spoken throughout the mountainous region of northern Pakistan. These languages exhibit rich, and typologically rare, coronal consonant inventories. They figure prominently in the discussion of retroflex consonant harmony in the following chapters.

1.1.3 Munda languages

The Munda language family consists of a small set of minority and mostly non-literary languages concentrated in eastern India, predominantly in the states of Orissa, Jharkhand and Chhattisgarh. The study of comparative Munda linguistics is still in relative infancy. Nevertheless, most accounts recognize two broad geographic groups, one north and one south, each with further possible sub-classifications. The main elements of the sub-classification scheme assumed in the present study are presented in Figure 4.

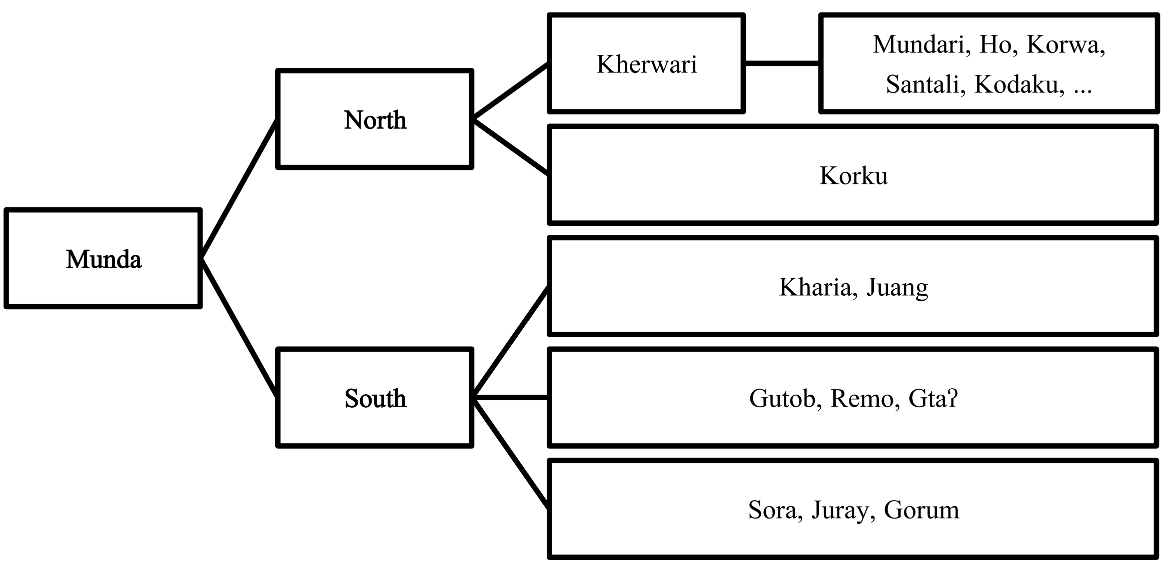


Figure 4 Sub-classification of Munda languages

1.1.4 Tibeto-Burman languages

The Tibeto-Burman family is large and extends well beyond the confines of South Asia, as defined here. The principal subgroups with representatives in South Asia are: (i) the greater "Himalayish" group, which subsumes most of the languages of northwest India and Nepal; and (ii) the Kuki-Chin-Naga group, the Jingpho-Konyak-Bodo group, Tani, Meitei, Mikir, and to a lesser extent, the Lolo-Burmese group, all of which are concentrated in northeast India.

The Tibeto-Burman languages do not figure prominently in the present study because most of them do not distinguish retroflex consonants. Those that do distinguish retroflex consonants occur primarily within the greater "Himalayish" group. The few Tibeto-Burman languages discussed in subsequent chapters all fall within this group. Thus, for reference purposes, only the sub-classification of the Himalayish group is presented in Figure 5.

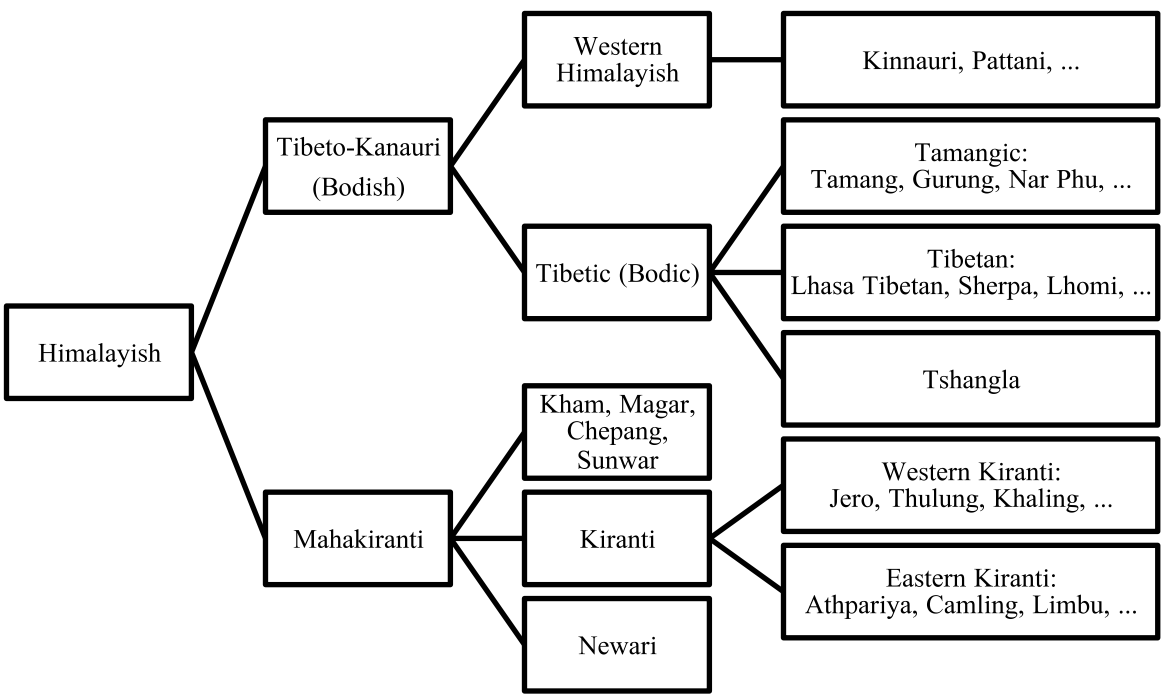


Figure 5 The Himalayish branch of Tibeto-Burman

The Tibetic languages (also referred to as Bodic, e.g., Thurgood, 2003) are of particular interest to the present study. Languages of this group that are discussed in subsequent chapters include Tibetan (§2.2.5), Lhomi (§2.3.2) and Sherpa (§3.5).

In summary, South Asia is a region of tremendous linguistic diversity, with as many as 539 languages representing four major language families. These language families share a long history of contact and convergence. As a result, South Asia now constitutes a linguistic area in which languages of different genetic stock share common areal traits. From a phonological point of view, the most prominent areal trait of South Asia is the widespread use of retroflexion in languages of all genetic stocks (Emeneau, 1956; Ramanujan & Masica, 1969; Bhat D. N., 1973; Masica, 1992). This makes South Asia an ideal region for the study of phonological phenomena pertaining to retroflexion, including retroflex consonant harmony. The following section provides some background on the nature of retroflexion.

1.2 Retroflexion and other coronal articulations

This dissertation is concerned primarily with phonological interactions involving retroflex segments. What is retroflexion? Traditionally, retroflexion is regarded as a place of articulation, primarily on the grounds that retroflex consonants can be distinguished for manner of articulation along the same lines as consonants at other places of articulation. Thus, languages can distinguish a retroflex series of consonants that includes plosives, affricates, fricatives, nasals, liquids and approximants parallel to other places of articulation.

Places of articulation are commonly described in terms of both active (i.e., moveable) and passive (i.e., immoveable) articulators. Active articulators include the lower lip for labial articulations, the tongue tip and/or blade for coronal articulations, the tongue body (or ‘dorsum’) for dorsal articulations, the tongue root for radical articulations and the glottis for glottal articulations. Of these, the coronal articulator is the most mobile. It is subdivided into laminal articulations formed with the tongue blade (or ‘lamina’) and apical articulations formed with the tongue tip (or ‘apex’). In phonetic descriptions, a distinction is sometimes made between apical articulations, formed with the tongue tip, and sub-apical articulations, formed with the underside of the tongue tip.

Passive articulators are the various regions along the upper and rear surface of the vocal tract that serve as sites of constriction targeted by the active articulators. At least nine target regions are generally recognized. They include the labial region (on the upper lip), the dental region (on or near the upper incisors), the alveolar region (on or in front of the alveolar ridge), the post-alveolar region (just behind the alveolar ridge), the palatal region (on the hard palate), the velar region (on the soft palate or ‘velum’), the uvular region (on the uvula), the pharyngeal region (on the pharyngeal wall below the uvula) and the epiglottal region (just above the larynx).

Retroflex consonants belong to the class of coronal articulations. Traditionally, they have been described as articulations produced by curling the tongue tip back toward the hard palate. This is reflected in the term “retroflex” which derives from Latin *retro* ‘backward’ + *flectere* ‘to bend’. Other common terms for this class of segments, particularly in the literature from the first half of the 20th century and earlier, include “cerebral” and “cacuminal”.

The following sections review the phonetic properties of retroflexion in terms of both articulation (§1.2.1) and acoustic/perceptual cues (§1.2.3). The relation of retroflex articulations to other coronal place articulations is discussed in §1.2.2.

1.2.1 Articulatory properties of retroflexion

Retroflex consonants exhibit a wide range of variation in terms of both active and passive articulators. A distinction is sometimes made between two degrees of retroflexion: a weaker form of apical post-alveolar retroflexion and a stronger form of sub-apical palatal retroflexion. Degree of retroflexion is sometimes said to be language dependent and possibly conditioned by genetic affiliation. For instance, apical post-alveolar retroflexion is said to be characteristic of Hindi and other Indo-Aryan languages while sub-apical palatal retroflexion is attributed to Tamil and other Dravidian languages (Ladefoged & Bhaskararao, 1983; Ladefoged & Maddieson, 1996). Representative examples of these two degrees of retroflexion are shown in Figure 6.

Figure 6 X-ray tracings of an apical retroflex plosive in Hindi and a sub-apical retroflex plosive in Tamil (adapted from Ladefoged & Maddieson, 1996, p. 27)

Although individual languages or language families may show a preference for one degree of retroflexion over another, variation in the degree of retroflexion has also been reported within individual languages of both Indo-Aryan and Dravidian affiliation (Reddy K. N., 1986; Dixit & Flege, 1991; Dart & Nihalani, 1999; Khatiwada, 2007). Language-internal variation can be conditioned by speaker, speech rate, manner of articulation, or vowel context. Retroflex consonants tend to be more sub-apical and palatal in the context of back vowels and more apical alveolar or post-alveolar in the context of front vowels. Significantly, no language is known to contrast two degrees of retroflexion (Ladefoged & Maddieson, 1996, p. 27).[7]

Hamann (2003) suggests that the two types of retroflexion in Figure 6 should be regarded as two extremes along a continuum of possible retroflex articulations. She identifies four articulatory properties that, taken together, define retroflexion. They include apicality, posteriority, sublingual cavity, and retraction. These properties are summarized in (2) and elaborated below.

(2) Articulatory properties of retroflexion (Hamann, 2003)

a. apicality: articulated with the tip or underside of the tongue

b. posteriority: articulated behind the alveolar ridge

c. sublingual cavity: articulated with a cavity beneath the tongue blade

d. retraction: displacement of the tongue dorsum towards the pharynx or velum

Hamann defines apicality as any articulation involving either the upper or lower side of the tongue tip. Thus, her definition subsumes both apical articulations proper (i.e., those articulated with the upper side of the tongue tip) and sub-apical articulations (i.e., those articulated with the underside of the tongue tip). In this sense, apicality is a consistent property

---

[7] Hamann (2003, pp. 66–68) suggests that the Dravidian language Toda distinguishes two retroflex fricatives, one apical post-alveolar and the other sub-apical and palatal. Hamann's source of information on Toda is Shalev, Ladefoged & Bhaskararao (1993) in which the corresponding fricatives are clearly described as apical alveolar and sub-apical postalveolar respectively (p. 111). Based on the description in the original source there is no reason to assume that the apical alveolar fricative is retroflex. It is not clear how or why Hamann arrived at different articulatory descriptions for these segments but the claim that both are retroflex is doubtful.

of retroflexion. Retroflex segments are always apical in the broad sense employed by Hamann; they are never laminal (i.e., articulated with the tongue blade).

As a general rule, retroflex segments also tend to be posterior, where posteriority is taken to mean 'articulated behind the alveolar ridge'. Posterior segments are distinguished from anterior segments, which are articulated on or in front of the alveolar ridge (i.e., in the denti-alveolar region). However, phonetic studies reveal a significant degree of variation in the passive place of articulation for retroflex segments. While they are predominantly posterior, retroflex segments can be articulated well into the anterior region in some cases, particularly in the context of front vowels (Dixit & Flege, 1991; Khatiwada, 2007). Moreover, retroflexion is characterized by a dynamic gesture that Hamann describes as 'flapping out'. The tongue tip is raised and drawn back for the onset of consonantal constriction but is released in a forward direction. As a result, the passive place of articulation is typically more posterior at the onset of constriction than at its release. The passive place of articulation at the release of a retroflex constriction is often close to the alveolar ridge and can even stray into the anterior region. Thus, of all the articulatory properties of retroflexion in (2), Hamann suggests that posteriority may be the only one that is non-essential. A segment having all of the articulatory properties in (2) except for posteriority can still be regarded as retroflex.

The raising of the tongue tip toward the post-alveolar region produces a large sublingual cavity, which Hamann identifies as another articulatory property of retroflexion. A sublingual cavity is also characteristic of other posterior coronal articulations including laminal post-alveolars, although the size of the cavity is largest for retroflex segments (Keating, 1991). Sublingual cavity contributes to the lowering of vocal tract resonance frequencies but does not represent either an active or passive articulator. Thus, it might be best regarded as a concomitant feature of the other articulatory properties in (2).

The fourth and final articulatory property of retroflexion identified by Hamann (2003) is retraction. Retraction refers to a displacement of the tongue dorsum toward the velum or pharynx, comparable to the gesture involved in secondary velarization or pharyngealization. Bhat (1974) suggests that retroflex consonants are preferentially retracted for ease of achieving an apical post-alveolar articulation but argues that retraction is neither unique to retroflexion nor a necessary component of it. Contra this claim, Hamann (2002; 2003) argues that retraction is a necessary but not sufficient criterion of retroflexion. According to Hamann's definition, all retroflex segments are inherently retracted but not all retracted segments are retroflex.

As the tongue tip is drawn up toward the post-alveolar region and the tongue dorsum is retracted toward the velum or pharynx, the middle of the tongue is lowered resulting in a concave tongue shape. Hamann regards lowering of the tongue middle (or concaving of the tongue dorsum) as a concomitant property of retraction when it is combined with apicality and posteriority (2003, p. 36).

In sum, retroflexion is classified as a coronal 'place' of articulation, and traditionally defined as the curling back of the tongue tip toward the palate. However, Hamann (2003) has argued that retroflexion can be decomposed into four articulatory properties: apicality, posteriority, sublingual cavity and retraction. Other articulatory properties of retroflexion, including the curling back of the tongue tip and the concaving of the tongue body, can be regarded as concomitant to these. The following section (§1.2.2) relates the articulatory properties of retroflexion to those of other coronal places of articulation.

1.2.2 Coronal places of articulation

Retroflex consonants belong to the class of coronal articulations. Traditional phonetics recognizes at least six passive places of articulation that are commonly classified as coronal because any constriction formed at those places is generally implemented with the tip or blade of the tongue. Coronal places include: inter-dental, dental, alveolar, palato-alveolar, alveolo-palatal and retroflex. However, not all passive place distinctions are phonologically relevant. For instance, few (if any) languages maintain contrast based solely on the distinction between dental and inter-dental place (Ladefoged & Maddieson, 1996, p. 20) or palato-alveolar and alveolo-palatal place (Hall, 1997b, p. 67). From a phonological point of view a contrast between four coronal places of articulation is probably maximal (Hall, 1997b, p. 88).[8] Four-way coronal place systems are typologically rare. Nevertheless, they are attested in Australian aboriginal languages (Dixon, 1980; 2002; Evans, 1995; Hamilton, 1996) and in some South Dravidian languages of India (see §2.1.2 and §2.2.2). Such systems typically include dental, alveolar and retroflex articulations along with some form of laminal post-alveolar articulation, whether palato-alveolar (closer to alveolar) or alveolo-palatal (closer to palatal). Within the South Asian tradition laminal post-alveolar coronals are often labeled loosely as "palatal". This convention is adopted here for convenience although it should be noted that such post-alveolar palatals are distinct from "true" palatals, which are arguably dorsal, not coronal (Ladefoged & Maddieson, 1996, pp. 32–33; Hall, 1997b). In the rare instances where a distinction must be

---

[8] Some languages distinguish more than four coronal fricatives. However, the additional contrasts are typically achieved through secondary articulations such as labialization, not by distinguishing additional coronal places (Hall 1997b, p. 94). One possible exception is the Dravidian language Toda, which is reported to have a five-way place contrast among coronal fricatives (Shalev et. al., 1993; Hall 1997b, p. 92).

maintained, the term palatal is reserved for the class of laminal post-alveolar coronals in the present study and the term "dorso-palatal" applied to all others.

Cross-linguistically, dental consonants tend to be laminal and alveolar consonants tend to be apical (Keating, 1991, p. 42; Ladefoged & Maddieson, 1996, pp. 20-21; Hall, 1997b, p. 42). Palatal and retroflex consonants are also laminal and apical, respectively. Thus, coronal consonants can be classified into sets of apical and laminal articulations. Within each of these classes a further distinction can be made between articulations formed with a more anterior place of constriction and those formed with a more posterior constriction, as shown in Table 2.

Table 2 Coronal place articulations

	Anterior	**Posterior**
Laminal	dental t̻	palatal ʧ
Apical	alveolar t̺	retroflex ʈ

The distinction between anterior and posterior consonants is often defined in relation to the alveolar ridge. Segments articulated on or in front of the alveolar ridge are classified as anterior while those articulated behind the alveolar ridge are classified as posterior. However, passive place of articulation can vary substantially for some coronals, especially those of the posterior class. The wide range of variation for retroflex articulations has already been noted in §1.2.1. Similar findings have been reported for palatals in some languages. For instance, Dart & Nihalani (1999) found that the so-called 'palatal' stops of Malayalam are actually laminal alveolar for most speakers. They conclude, "the traditional place label *palatal* or *palatoalveolar* does not correspond to the phonetic reality for this segment" (p. 135). Thus, passive place of articulation is not the most reliable factor for distinguishing coronal articulations and it is best to take the labels 'anterior' and 'posterior' as referring to relative positions, not absolute points or discrete regions of the palate.

In terms of active articulators, posterior coronals may also be distinguished from anterior coronals by activity of the tongue dorsum. For instance, palatals are typically articulated with a fronted tongue body and raised tongue middle resulting in a convex or "domed" tongue shape. This tongue shape is comparable to that of high front vowels and accounts for the natural affinity between palatal consonants and high front vowels cross-linguistically. Similarly, retroflex consonants tend to be articulated with a retracted dorsum and lowered tongue middle resulting in a concave tongue shape. This tongue body gesture accounts for the natural affinity between retroflex consonants and low and/or back vowels, as well as the mutual incompatibility of retroflexion and palatalization (Hamann, 2003). Thus, in terms of active articulators it might be useful to think of posterior coronals as those with strong inherent dorsal components and anterior coronals as those without (Arsenault, 2008; 2009b).

In sum, phonological systems contrast up to four coronal places of articulation. They can be labeled loosely as dental, alveolar, retroflex and palatal. Each of these can be classified according to one of two active articulators (i.e., laminal vs. apical) and one of two passive target regions (i.e., anterior vs. posterior). However, passive place of articulation is often subject to variation and target regions can overlap with each other. As a result, they must be understood in relative terms. Alternatively, they may be conditioned by the presence or absence of inherent dorsal articulations.

1.2.3 Acoustic and perceptual properties of retroflexion

The preceding sections reviewed the articulatory properties of retroflexion (§1.2.1) and other coronal places of articulation (§1.2.2). The perception of acoustic cues pertaining to place of articulation has also been argued to play an important role in the phonological patterning of retroflex consonants and other places of articulation (e.g., Hamilton, 1996; Steriade, 2001; Hamann, 2003). This section reviews the most relevant generalizations concerning the acoustic cues and perceptual properties of retroflexion and other coronal places.

An acoustic cue can be defined as information in the acoustic signal that enables the listener to apprehend a phonological contrast (Wright, 2004, p. 36). Cues to place of articulation can be found primarily in formant patterns and spectral shape, though other factors, such as the duration of constriction or voice onset time (VOT), can also play a role in some cases. Cues can be internal to a consonant (i.e., during the period of constriction) or external, in which case they are found in the period of transition between a consonant and an adjacent segment, most notably an adjacent vowel. Transitions from a consonant into a vowel are commonly known as CV transitions. Transitions from a vowel into a consonant are known as VC transitions.

Consonants of all manners exhibit external place cues in the form of formant transitions into or out of adjacent vowels. Thus, CV and VC formant transitions are probably the most frequently cited cues to place of articulation. Oral stops are characterized by a period of complete obstruction to airflow. Thus, they lack internal place cues altogether. Their place of articulation must be perceived entirely on the basis of external cues, which include CV and VC formant transitions as well as spectral information in the release burst during the CV transition. Other manners of articulation exhibit internal place cues in addition to external formant transitions. The place of nasals can be partly distinguished on the basis of internal resonance patterns. Fricatives exhibit different internal spectral noise patterns depending on their place of articulation. Liquids and approximants maintain some internal formant structure indicative of their place of articulation. The most commonly cited sources of consonantal place cues are summarized in Table 3, following Jun (2004).

Table 3 Sources of consonant place cues (adapted from Jun, 2004, p. 60)

	Cue type	
Segment type	**Internal**	**External**
stops	none	CV, VC formant transitions, release burst
nasals	nasal resonance	CV, VC formant transitions
fricatives	frication noise	CV, VC formant transitions
liquids and glides	formant structure	CV, VC formant transitions

Most sources agree that the single most consistent cue to retroflexion is a lowering of the upper formants, most notably F3 (Ladefoged & Maddieson, 1996, pp. 27-28; Hamilton, 1996, pp. 47-48; Hamann, 2003, p. 59ff.). Abstracting away from language-specific and contextual variation, Hamann (2003) summarizes the relative height of F2 and F3 formant frequencies for coronal places of articulation as shown in Table 4.

Table 4 Relative formant frequencies for coronal places of articulation (adapted from Hamann, 2003, p. 63)

	t̻	t̺	ʈ	ʧ
F3	mid	mid	lowest	highest
F2	high	high	high	highest

The generalizations in Table 4 indicate that all coronal articulations tend to exhibit a relatively high F2 in comparison to non-coronal consonants. Thus, F2 may serve as a cue to coronal place more broadly but it is not necessarily a reliable indicator of place distinctions within the coronal class. Laminal postalveolar ("palatal") articulations are a possible exception. They are often reported to have a distinctively high F2; one comparable to that of the cardinal high front vowel [i] (Hamilton, 1996, p. 47). Otherwise, the F2 trajectories of most coronals are overlapping and largely dependent on vocalic context.

Retroflex consonants exhibit the lowest F3 values of all places in Table 4. A low F3 consistently distinguishes retroflex consonants from all other places of articulation, coronal or non-coronal. This is manifested most clearly in VC transitions where a dramatic lowering of F3 can be seen in spectrograms for retroflex consonants of all manners. Retroflex fricatives are also characterized by a lower frequency noise pattern relative to other coronal fricatives. Retroflex articulations may also be characterized by shorter closure durations and shorter VOT relative to other coronals (Hamann, 2003, p. 69).

Notice that lamino-dental and apico-alveolar articulations have overlapping F2 and F3 trajectories in Table 4. No doubt, this accounts for the fact that contrast between these two articulations is rare cross-linguistically (Maddieson, 1984, pp. 31-32). Where they are distinguished, some studies suggest that perceptual cues may be found in their release burst. Laminal articulations are said to have a longer, noisier release (and hence, a greater propensity for affrication) while apical articulations have a cleaner, more abrupt release (Hamilton, 1996, pp. 50-51; Hamann, 2003, p. 56).

Acoustic cues are not equally prominent in every context. For instance, while place cues can be found in both CV and VC transitions, a growing body of evidence suggests that CV transitions provide more prominent place cues than VC transitions for most places of articulation (see Steriade 2001 and Jun 2004 and the sources cited therein). In experimental conditions, listeners attend to cues in CV transitions over those in VC transitions to identify place of articulation. However, retroflex articulations constitute an important exception to this trend. Whereas most places of articulation are best cued in CV contexts (over VC), retroflexion is best cued in VC contexts (over CV) (Ladefoged & Maddieson, 1996, p. 28; Hamilton, 1996; Steriade, 2001; Hamann, 2003).

The explanation for this reversal of contextual prominence stems from one of the unique articulatory properties of retroflexion, namely the dynamic “flapping out” gesture described in §1.2.1. Recall that the tongue tip is drawn back toward the postalveolar region for the onset of retroflexion but moves forward to a more anterior constriction at the point of release. The initial retraction of the tongue tip toward the postalveolar region results in a dramatic lowering of F3 in the VC transition. This is often perceived as a contextual “r-colouring” on the preceding vowel. However, the more anterior release of retroflex consonants results in a less distinctive CV transition. In fact, some studies suggest that the CV transitions of retroflex consonants are almost indistinguishable from those of dental and alveolar consonants (Hamilton, 1996, p. 48; Steriade, 2001, p. 225).

In sum, retroflexion is a type of coronal articulation that is apical and (typically) posterior. Retroflex segments are also characterized by a retracted tongue dorsum, concave tongue shape and a large sub-lingual cavity. The most reliable acoustic cue to retroflexion is a low F3, which is most prominent in VC formant transitions. In this respect, retroflexion differs from most other places of articulation, for which acoustic cues are more perceptually salient in CV contexts. The asymmetry between the contextual prominence of retroflex cues and other place cues is important. As we will see, it has been cited to explain phonotactic constraints on retroflexion (§2.3) and the unique behaviour of retroflex segments in patterns of local place assimilation (§2.4).

1.3 Consonant harmony

As already noted, this dissertation is concerned with phonological interactions involving retroflex segments. More specifically, it is concerned with those phonological interactions that can be classified as cases of consonant harmony. A definition of consonant harmony is provided in §1.3.1. Section §1.3.2 introduces the typological parameters that are relevant to the study of consonant harmony, and section §1.3.3 provides a brief overview of those typological properties that are most characteristic of consonant harmony systems *vis-à-vis* other types of assimilation. The discussion throughout this section draws heavily on work by Hansson (2001; 2010), Rose & Walker (2004) and Rose (2011).

1.3.1 Definition of consonant harmony

Consonant harmony is a form of *long-distance assimilation* between consonants. Rose (2011) provides the definition shown in (3).

(3) Definition of consonant harmony (Rose, 2011, p. 1811)

Assimilation for an articulatory or acoustic property between two or more non-adjacent consonants, where intervening segments are not noticeably affected by the assimilating property.

Apart from the fact that consonant harmony involves a phonological interaction between two or more consonants, the definition in (3) includes two other crucial components: (i) consonant harmony is *assimilatory* in nature, and (ii) it is a *long-distance* or *non-local* form of assimilation. Each of these points is elaborated in turn below.

To begin with, consonant harmony is a form of *assimilation* between consonants. Assimilation can be understood as a process in which one segment becomes more like another by taking on some property (or properties) of that other segment, or as a co-occurrence restriction requiring a set of segments within some domain to be identical with respect to a given property (or properties). Thus, assimilation can manifest itself in the form of dynamic morphological alternations or static morpheme structure constraints (MSCs).

In the case of alternations, the properties of a segment in one morpheme vary in accordance with those of a segment in another morpheme. This results in observable synchronic variation in the realization of a morpheme depending on its context. For example, in Benchnon (a.k.a., Gimira; Afro-Asiatic), the causative suffix /-s/ surfaces as retroflex [-ʂ] after roots containing a retroflex sibilant, palatal [-ʃ] after roots containing a palatal sibilant, and alveolar [-s] otherwise. Representative examples are shown in (4). In this case, the sibilant in the suffix takes on the place properties of the sibilant in the root. This results in synchronic alternations in the causative morpheme, which surfaces variably as alveolar, retroflex or palatal.

(4) Benchnon consonant harmony: Alternations in causative /-s/ (Hansson, 2010, p. 52)

s^jap-s-	'make wet'
ʃir-ʃ-	'bring near'
s̺up-s̺-	'make soft'

MSCs, by comparison, are static co-occurrence restrictions on segments within a morpheme. Although they do not result in observable variation, MSCs are considered *assimilatory* when they require segments to be identical with respect to certain properties, and *dissimilatory* when they require them to differ with respect to those properties. For example, in addition to the alternations shown above, Benchnon also restricts the co-occurrence of sibilants within morphemes. If there are two sibilants within a morpheme, then they must agree with respect to place of articulation, as shown in (5).

(5) Benchnon consonant harmony: Root-internal MSC (Hansson, 2010, p. 52)

a. *Well-formed roots*

sis	‘fir tree’
ʦ’oʦ’-	‘centre’
ʃaʃkn	‘green tree-snake’
ʧiʃkn	‘bile’
s̢ets̢’	‘type of cabbage’
ts̢’onts̢’	‘fill (tr.)’

b. *Prohibited root-internal sequences*

*s…ʃ, *s…s̢, *ʃ…s, *ʃ…s̢, etc.

*s…ʧ, *s…ts̢, *ʃ…ʦ, *ʃ…ts̢, etc.

*ʦ…ʃ, *ʦ…s̢, *ʧ…s, *ʧ…s̢, etc.

*ʦ…ʧ, *ʦ…ts̢, *ʧ…ʦ, *ʧ…ts̢, etc. (etc.)

Some languages exhibit assimilation only in the form of static MSCs. Others, like Benchnon, exhibit assimilation in both MSCs and alternations. The two forms of assimilation are fundamentally the same; they differ only with respect to the domain over which an assimilatory co-occurrence restriction applies. In the case of MSCs, the restriction is limited to the domain of the morpheme and does not extend across morpheme boundaries to produce alternations. In the case of alternations, the restriction extends to a larger domain that includes morphologically complex words/stems consisting of roots plus affixes.

In any assimilation pattern, a segment that extends its properties to other segments is commonly known as a *trigger* of assimilation. Any segment that derives (some or all of) its properties from the trigger is a *target* of assimilation. In the case of the Benchnon causative suffix in (4), the alveolar sibilant /-s/ is the target of assimilation, while palatal and retroflex sibilants, such as /ʃ/ and /ʂ/, are the triggers. In the absence of alternations, it can be difficult or even impossible to identify triggers and targets reliably. However, even in such cases the triggers and targets of assimilation can often be established on the basis of historical-comparative evidence. Root-internal MSCs are often the product of the same assimilatory processes that drive alternations, only applied diachronically as opposed to synchronically. Where alternations are lacking, cognates from a historically prior form of the language, or from closely related languages or dialects, often reveal patterns of diachronic assimilation in which some segments have served as triggers while others have served as targets.

The second crucial component of the definition in (3) is that consonant harmony refers exclusively to *long-distance* assimilation. Broadly speaking, assimilation can be *local*, in which case it holds between consonants that are adjacent in the phonological string, or it can be *non-local* or *long-distance*, in which case it holds between consonants that are non-adjacent. By definition, the trigger and target of consonant harmony must be non-adjacent in the phonological string. At the very least, they must be separated by a vowel, although they may be separated by a longer string of segments consisting of (any number of) vowels and consonants, as in some of the Benchnon examples in (4) and (5).

The definition of consonant harmony in (3) includes an important criterion that bears on the issue of locality: intervening segments between the trigger and target "are not noticeably affected by the assimilating property." This criterion is intended to exclude apparent cases of long-distance assimilation that are really the product of iterative serial applications of local assimilation. The distinction between long-distance assimilation and the serial application of local assimilation is represented schematically in (6).

(6) Local vs. long-distance assimilation

a. *serial local assimilation*

$A_{[F]}\ B_{[F]}\ C_{[F]}$

b. *long-distance assimilation*

$A_{[F]}\ B\ C_{[F]}$

In a string of segments, ABC, where A and C are non-adjacent consonants, assimilation of some feature [F] constitutes a case of consonant harmony only if [F] is extended directly from C to A (or A to C) without regard for the intervening segment B. This is represented schematically in (6)(b). The scenario in (6)(a) is similar in that [F] also extends from C to A (or A to C). However, in this case, it does so indirectly through the intervening segment B, which is also targeted by the assimilating feature. Although (6)(a) has the appearance of a long-distance interaction, it is arguably the product of a strictly local interaction applied iteratively in a serial manner: B assimilates the feature locally from C; and A, in turn, assimilates it locally from B. Since there is no direct interaction between C and A, cases of assimilation that fit the pattern in (6)(a) are not included in the definition of consonant harmony adopted here.

Within the phonological literature, the term ‘harmony’ is sometimes applied to assimilatory patterns that are not necessarily long-distance in the narrow sense adopted here. For instance, the term ‘nasal harmony’ is often applied to a pattern of assimilation like that in (6)(a), where nasalization targets a contiguous string of segments (e.g., Piggott, 1992; Walker, 1998). In light of this, Rose & Walker (2004) propose the term Long Distance Consonant Agreement (LDCA) as an alternative to ‘consonant harmony’ for patterns that fit the description of (6)(b). Throughout the dissertation, I retain the more traditional term 'consonant harmony' with the qualification that the term, as used here, is synonymous with Rose & Walker's LDCA, and refers exclusively to assimilation patterns like that in (6)(b).

In sum, consonant harmony can be defined as long-distance assimilation between consonants. It can be manifested in the form of alternations or static co-occurrence restrictions on consonants within morphemes (MSCs). The following section reviews some additional parameters relevant to the study of consonant harmony systems.

1.3.2 Parameters of assimilation

We have seen that assimilation can be local or non-local, and that it can apply over various domains (e.g., morpheme-internally or across morpheme boundaries). These represent two parameters by which assimilation patterns can be described and classified. A more extensive list of parameters relevant to the study of consonant harmony and other assimilation patterns is presented in (7). Each of these parameters is discussed in turn below.

(7) Some parameters of assimilation

a. Locality: Are the interacting segments adjacent or non-adjacent?

b. Dominance: Which segments/properties serve as triggers of assimilation, and which as targets?

c. Domain: Is assimilation restricted to roots or does it extend to larger domains?

d. Direction: Is assimilation regressive, progressive, or stem-controlled?

e. Transparency/Opacity: Are intervening segments transparent to assimilation or do they potentially block assimilation?

The parameters of locality, in (7)(a), and domain, in (7)(c), have already been discussed in the preceding section. By definition, consonant harmony is non-local, and the domain of harmony can be the morpheme (in which case harmony is manifest only as a static co-occurrence restriction on roots) or some larger domain that extends across morpheme boundaries (in which case it may produce alternations).

Another parameter is the assimilating property itself and the class of segments that interact as triggers and targets (7)(b). Cross-linguistically, the most frequent type of consonant harmony is coronal harmony, which involves the long-distance assimilation of coronal place features, including retroflexion, as in the case of Benchnon (Gimira) cited above (see (4) and (5)). However, contrary to earlier assumptions (e.g., Gafos, 1999), consonant harmony is not limited to cases involving coronal features. Recent cross-linguistic surveys by Hansson (2001; 2010) and Rose & Walker (2004) have revealed that consonant harmony comes in many varieties. It can involve the assimilation of coronal features, dorsal features, laryngeal features, liquid features, nasality, and possibly even stricture features and secondary articulations. The present study is concerned primarily with the long-distance assimilation of retroflexion, although assimilation of other features is also discussed wherever it is relevant.

The class of triggers and targets can be symmetrical or asymmetrical. In a symmetrical system, any segment that participates in the assimilation pattern can serve as either a trigger or a target. For instance, in some languages with coronal sibilant harmony involving denti-alveolar and postalveolar sibilants, postalveolars can trigger assimilation in denti-alveolars and *vice versa* (/s…ʃ/ → [ʃ…ʃ]; /ʃ…s/ → [s…s]; e.g., Chumashan languages; Hansson, 2010, pp. 44–45). In asymmetrical systems, some segment types (or features/properties) serve exclusively as triggers, while others serve exclusively as targets. For example, in other sibilant harmony systems, denti-alveolar sibilants serve as targets of assimilation, but not triggers, while postalveolar sibilants serve as triggers, but not targets (/ʃ…s/ → [ʃ…ʃ]; but /s…ʃ/ ↛ [s…s]; e.g., Benchnon; Hansson, 2010, p. 53). Segments (or features/properties) that serve only as triggers are said to be *dominant*, while those that serve only as targets are said to be *recessive* (Baković, 2000; Hansson, 2010).

Another important parameter of assimilation is the direction of assimilation (7)(d), which can be regressive, progressive, or stem-controlled. In regressive assimilation, the trigger follows the target (A…B → B…B). In progressive assimilation, the trigger precedes the target (A…B → A…A). In the case of stem-control, the direction of assimilation is determined by morphological factors: segments in affixes assimilate to segments in roots/stems. This can result in regressive assimilation if a language has prefixes, progressive assimilation if it has suffixes (cf. Benchnon examples in (4)) or bi-directional assimilation if it has both.

Finally, segments that intervene between the trigger and target of assimilation can be classified as *transparent* or *opaque* (7)(e). Intervening segments are said to be transparent when they do not inhibit the assimilating property from reaching potential targets within the domain of harmony. They are said to be opaque when they *block* the assimilating property and prevent it from reaching potential targets within the harmony domain. This parameter is discussed further, with examples, in §1.3.3.2, below.

Assimilation patterns of any type can be described and classified in terms of the parameters outlined above. However, not all of the properties cited here are equally attested in consonant harmony systems. The following section reviews some of the most common typological properties of consonant harmony systems *vis-à-vis* other patterns of assimilation.

1.3.3 Typological properties of consonant harmony systems

Cross-linguistic surveys identify at least three typological properties that are characteristic of consonant harmony systems *vis-à-vis* other patterns of assimilation, which include local consonant assimilation, vowel harmony, and what is sometimes referred to as vowel-consonant harmony. These properties are: (i) the similarity of interacting segments (§1.3.3.1); (ii) the transparency of segments intervening between the trigger and target of assimilation (§1.3.3.2); and (iii) an inherent bias toward regressive assimilation (§1.3.3.3). Each of these properties is described and exemplified below.

1.3.3.1 Similarity of interacting segments

Similarity effects are perhaps the most prominent typological property of consonant harmony systems in relation to local assimilation. In consonant harmony, the class of interacting segments is typically limited to a small set of highly similar segments. That is, assimilation for some feature [F] holds only between those segments in the system that are similar to one another by virtue of sharing a large number of other phonological features. This is not always the case in other patterns of assimilation where the class of interacting segments is typically much larger and often unconstrained by similarity.

For example, local voicing assimilation often holds between all obstruents in a language, regardless of stricture. English is representative of this pattern. In English plural and past tense formations, obstruent clusters agree in voicing regardless of whether they agree in stricture or not. Thus, we find ca[ts] ‘cats’ and do[gz] ‘dogs’, with voicing assimilation between plosives and sibilants, in addition to wal[kt] 'walked' and jo[gd] 'jogged', with voicing assimilation between two plosives. In contrast to this, voicing harmony is typically restricted to a subset of highly similar obstruents (Rose, 2011, p. 1826). Consider the example of Tamajaq Tuareg (Berber) in (8).

(8) Tamajaq Tuareg: voicing harmony between sibilants (Hansson, 2010, p. 114)

	Base		*Causative*	
a.	əlməd	'learn, study'	s-əlməd	'teach, inform'
	busu	'be injured'	s-əbːusu	'injure'
b.	əntəz	'pull out, extract'	z-əntəz	'cause to extract'
	əbzəg	'be mad, panic'	z-əbzəg	'drive mad, cause to panic'
	guləz	'be left, remain'	z-əgːuləz	'cause to remain'

In Tamajaq Tuareg, voicing harmony produces alternations in the causative prefix /s-/, which surfaces as [z-] before roots containing a voiced sibilant, and as [s-] otherwise.[9] Significantly, voicing harmony holds only between sibilants. It is not triggered by other voiced consonants, including voiced stops and sonorants. For instance, the causative prefix is voiced before a root containing [z], as in [z-əntəz] 'cause to extract', but not before a root containing [d], as in [s-əlməd] 'teach, inform'.

Similarity effects of this kind are found in all types of consonant harmony. For instance, laryngeal harmony is often dependent on similarity of place and manner; nasal consonant

---

[9] This is a simplification. The causative prefix of Tamajaq Tuareg is also subject to coronal harmony. As a result, it has two additional allophones: [ʃ-] and [ʒ-]. This detail is omitted here because it is not directly relevant to the point at hand. See Hansson (2010, p. 114) for details.

harmony is often dependent on agreement in voicing and/or sonorancy; coronal place harmony is often dependent on manner in terms of the obstruent vs. sonorant distinction and/or the sibilant vs. non-sibilant distinction; and liquid harmony, which involves assimilation of the feature responsible for the lateral vs. non-lateral distinction, typically applies only to the class of liquids (including laterals and rhotics) or possibly to the class of sonorants, but not, for instance, to the class of all coronals. In sum, similarity effects are highly characteristic of consonant harmony systems but not necessarily characteristic of local assimilation patterns.[10]

1.3.3.2 Transparency of intervening segments

Another prominent characteristic of consonant harmony systems is the neutrality or transparency of segments that intervene between the trigger and target of assimilation. Intervening segments rarely block assimilation in consonant harmony systems. In contrast to this generalization, blocking effects are routinely observed in vowel and vowel-consonant harmony systems. By way of example, consider the pattern of nasal consonant harmony in Yaka (Bantu), shown in (9).

---

[10] So-called 'parasitic' vowel harmony systems may also show sensitivity to similarity. For example, rounding harmony in Yowlumne (a.k.a. Yawelmani) applies only between vowels that agree in height (Kuroda, 1967). It remains unclear whether cases of this type are the product of local or long-distance assimilation (cf. Rose & Walker, 2004, pp. 490, 520).

(9) Yaka: Nasal consonant harmony (Hansson, 2010, p. 86)

a.	-són-ene	'color'	cf. -sól-ele	'deforest'
	-kém-ene	'moan'	cf. -kéb-ele	'be careful'
	-ján-ini	'cry out in pain'	cf. -jád-idi	'spread'
	-ʦúm-ini	'sew'	cf. -ʦúb-idi	'wander'
b.	-mák-ini	'climb'		
	-ɲéːk-ene	'bend down'		
	-nútúk-ini	'bow'		
	-míːtuk-ini	'sulk'		

In (9), the Yaka perfective suffix is realized as [-idi] or [-ele] in stems with oral consonants, but surfaces instead as [-ini] or [-ene] in stems containing nasal consonants.[11] The triggering nasal in the root and the target segment in the suffix are not adjacent. Assimilation extends across intervening vowels, as shown in (9)(a), and over longer spans consisting of vowels and consonants, as shown in (9)(b). All intervening segments, whether consonant or vowel, are neutral and transparent with respect to nasalization. Intervening vowels and consonants are not nasalized. Nor do they prevent nasalization from reaching the target segment in the suffix.

Johore Malay (Austronesian) exhibits a pattern of nasal assimilation with a very different set of properties. Patterns of this type are sometimes referred to as vowel-consonant

[11] The alternation between [i]~[e] is the result of vowel harmony. Variation between [l]~[d] is allophonic and is conditioned by segmental context; [d] occurs before [i] (and in NC clusters), [l] occurs elsewhere.

harmony. Representative examples are shown in (10). Following Hansson (2010), triggering nasals are shown in boldface, and all targeted segments are underlined.

(10) Johore Malay: Nasal consonant-vowel assimilation (Hansson, 2010, p. 85, 141)

mãj̃ãŋ	‘stalk (palm)’
mə̃**n**ãw̃ãn	‘to capture (active)’
mãʔ̃ãp	‘pardon’
pə**n**ə̃**ŋ**ãh̃ãn	‘central focus’
pə**ŋ**ãw̃ãsan	‘supervision’
mə̃ratappi	‘to cause to cry’

In Johore Malay, nasalization is triggered by a nasal consonant and targets a contiguous string of segments to the right of that consonant. Unlike Yaka, nasalization in Johore Malay is clearly a local phenomenon; it does not skip over any segment. Potential targets include vowels, glides and glottal consonants. Liquids and obstruents are not targeted. Instead of exhibiting transparency, all non-target segments block the nasal feature from reaching other potential targets to their right. Thus, in [pə**ŋ**ãw̃ãsan] ‘supervision’, nasalization targets the first two vowels and the labial glide to the right of the trigger ([ŋ]) but is prevented from reaching the final vowel because of the intervening [s].

In sum, consonant harmony systems operate over non-contiguous segments. Intervening segments in consonant harmony domains tend to exhibit transparency effects like that in Yaka, but rarely exhibit blocking effects.[12] In contrast to this, blocking effects are commonly observed in vowel-consonant harmony systems, like that in Johore Malay, which arguably involve iterative local assimilations.

1.3.3.3 Bias toward regressive assimilation

A third typological property of consonant harmony systems is their bias toward regressive assimilation. It is possible to find examples of consonant harmony involving regressive, progressive and bi-directional assimilation. However, Hansson (2001; 2010) has argued that all cases involving progressive or bi-directional assimilation are the result of stem control. That is, progressive and bi-directional assimilation arise in consonant harmony only when segments in affixes assimilate to segments in roots/stems. Wherever morphological constituency is not a factor, assimilation is almost exclusively anticipatory in nature and regressive in direction.

This is not necessarily the case in other patterns of assimilation where direction is often conditioned entirely by independent factors. For example, the direction of place assimilation in consonant clusters is often attributed to the distribution of perceptual cues. This accounts for a trend toward regressive assimilation when major place is involved (Jun, 2004), and progressive assimilation when retroflexion is involved (Steriade, 2001) (cf. §1.2.3 and §2.4). Similarly, studies of vowel harmony have argued that direction of assimilation is conditioned entirely by either stem control, in which segments in affixes assimilate to segments in stems, or dominance

---

[12] Exceptional cases of blocking in consonant harmony include: (Kinya)Rwanda (Bantu), where coronal sibilant harmony is blocked by intervening coronal consonants (Walker & Mpiranya, 2005; Walker, Byrd, & Mpiranya, 2008), and the Imdlawn and Agadir dialects of Tashlhiyt (Berber), where voicing harmony is blocked by intervening voiceless obstruents. Coronal sibilant harmony may also be blocked by intervening palatal [ç] in Tamazight (Berber). See Hansson (2010, pp. 166–175) for discussion of all these cases.

effects, in which certain features or segment types consistently trigger assimilation in others, irrespective of direction (Aoki, 1968; Baković, 2000).[13]

There may be exceptions to Hansson's (2001; 2010) generalization that consonant harmony is always regressive when it is not conditioned by stem control. For instance, some cases of nasal consonant harmony in Bantu appear to be progressive morpheme-internally (Rose & Walker, 2004, p. 490; cf. Hansson, 2010, pp. 88, 156–157, 289ff). Nevertheless, a strong trend or bias toward regressive assimilation appears to be a distinguishing mark of consonant harmony systems, one that is not necessarily shared by other assimilation patterns.

In sum, consonant harmony systems are characterized by at least three typological properties: (i) similarity effects; (ii) transparency effects; and (iii) regressive direction. These properties may not be absolute requirements of all consonant harmony systems, but at the very least they represent a clear typological trend. This trend stands out when consonant harmony systems are compared with other assimilation patterns, including local consonant assimilation, vowel harmony, and vowel-consonant harmony, where different properties and trends are evident. The typological properties of consonant harmony systems raise intriguing issues for phonological theories. Two of these issues are briefly introduced in the following section.

---

[13] Mahanta (2007) reports a case of regressive vowel harmony in Asamiya (Indo-Aryan) that cannot be reduced to stem control or dominant-recessive relations. If this is so, then it is possible that vowel harmony systems also default to regressive assimilation when other factors are not at play. However, it is not possible to make a reliable generalization on the basis of a single example. This is clearly a topic that requires further investigation.

1.3.4 Theoretical issues

Consonant harmony systems present many intriguing issues for phonological theory. Of these, only two are discussed in any depth in the present study. The first issue concerns the mechanism(s) that drive assimilation in consonant harmony. The second issue concerns the role of similarity in conditioning consonant harmony and the criteria by which languages evaluate similarity between interacting segments in consonant harmony systems. These two issues are introduced briefly here in anticipation of a more thorough discussion in later chapters.

Accounts of consonant harmony in non-linear phonology have operated under the null hypothesis that all assimilation, whether local or non-local, is accomplished by means of a single mechanism, which can be described as *feature spreading* or *gesture extension*. Feature spreading is essentially a local phenomenon: a feature or gesture is extended from one segment to another when the two segments are, in some sense, adjacent in the phonological string. Under this hypothesis, the non-local nature of consonant harmony can be explained by redefining locality relative to some structural unit independent of the segment; for instance, autosegmental tiers (e.g., Shaw, 1991). Alternatively, it has been argued that the non-local nature of consonant harmony might be a perceptual illusion. From this point of view, consonant harmony is accomplished by extending an articulatory gesture over a contiguous span of segments. All segments in the span are permeated by the spreading gesture, without exception, but some may appear to be transparent if the gesture has little or no audible effect on them. This is said to be particularly true in the case of coronal harmonies, including those that involve retroflexion, because coronal features/gestures are often irrelevant for vowels and non-coronal consonants (Gafos, 1999; Ní Chiosáin & Padgett, 1997; 2001).

Recent observations about the typological properties of consonant harmony systems, like those reviewed in §1.3.3, have prompted some researchers to question the hypothesis that all assimilation is the product of a single mechanism. They argue that local assimilation is the product of feature spreading while consonant harmony, which exhibits unique typological properties, is driven by a different mechanism, which they term feature *agreement*. Unlike spreading, feature agreement is genuinely non-local and induced by similarity. The proponents of this view suggest that agreement is grounded in the psycholinguistic domain of speech planning. When speakers are producing one segment, they are simultaneously planning the implementation of subsequent segments. Interference can occur if a segment that is being planned is highly similar to a segment that is being produced. One impinges on the other, resulting in familiar slips of the tongue, such as *shubjects show* for *subjects show*. If the same psycholinguistic mechanism responsible for errors of this type also lies behind consonant harmony, then it might explain the non-local nature of assimilation, with transparency of intervening segments, in addition to the similarity effects and the trend toward anticipatory (regressive) assimilation, all of which are characteristic of consonant harmony systems (Hansson, 2001; 2010; Rose & Walker, 2004).

A second theoretical issue surrounding consonant harmony concerns the criteria by which languages determine the set of interacting segments. While it is often claimed that interacting segments in consonant harmony systems are constrained by similarity, it is not clear how similarity is evaluated. Similarity may be evaluated over abstract representations made up of phonological features (Pierrehumbert, 1993), natural classes computed on the basis of phonological features (Frisch, Pierrehumbert, & Broe, 2004), or acoustic/perceptual properties (Gallagher, 2010; 2012). The features relevant to the evaluation of similarity in a language may be constrained by contrast (Mackenzie, 2009; 2011), or contrast may not be a reliable determining factor (Rose & Walker, 2004).

In sum, two major theoretical issues surrounding consonant harmony are: (i) the mechanism or mechanisms responsible for consonant harmony, whether spreading or agreement; and (ii) the criteria by which languages evaluate similarity and determine the set of interacting segments in consonant harmony systems. Both of these issues are taken up in later chapters and discussed in light of evidence from retroflex consonant harmony in South Asia.

1.4 Statistical methods

In languages where consonant harmony is strictly morpheme-internal, it is manifested only as a static co-occurrence restriction on consonants in lexical roots. In the absence of alternations, co-occurrence restrictions on non-adjacent consonants can be difficult to observe, as there is little to call attention to them. In recent years, statistical methods have become a popular means of examining long-distance co-occurrence restrictions on consonants, whether assimilatory or dissimilatory (e.g., Frisch, Pierrehumbert, & Broe, 2004; Kawahara, Ono, & Sudo, 2006; Coetzee & Pater, 2008; Gallagher & Coon, 2009). These methods have proven useful in highlighting categorical and gradient co-occurrence patterns that might otherwise go unnoticed. This section provides a brief introduction to the general method employed in much of the literature. This method is adopted in subsequent chapters to examine co-occurrence restrictions on retroflex and other coronal consonants in South Asian languages.

The statistical method employed in much of the literature, and also adopted in the present study, 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 roots/words 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 under random co-occurrence of consonants in the data set.[14] 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, the pair occurs as expected under random co-occurrence of the consonants 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 thus 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 thus disfavoured or avoided to some degree. Finally, an O/E ratio of 0.0 indicates categorical absence of the pair, 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'.

The statistical method can be illustrated with a few simple examples. Let us assume a hypothetical Language X. This language has a coronal consonant inventory consisting of dental

---

[14] The expected value for any C_1-C_2 sequence is calculated as the probability of C_1 occuring as the first member of the sequence, multiplied by the probability of C_2 occurring as the second member of the sequence, multiplied by the total number of C_1-C_2 sequences in the data set. The probability of a consonant occurring in a given position is calculated as the total number of sequences with that consonant in that position divided by the total number of sequences in the data set. For example, if we have a data set consisting of 500 C_1-C_2 sequences and the consonant /t/ appears as the first member in 40 of those sequences, then the probability of /t/ as the first member is calculated as $40 \div 500 = 0.08$. If /d/ appears as the second member in 90 sequences, then the probability of /d/ as the second member is calculated as $90 \div 500 = 0.18$. The expected value for the sequence /t…d/ is then calculated as $0.08 \times 0.18 \times 500 = 7.2$.

and retroflex stops: /t, ʈ/. In roots containing two non-adjacent coronal consonants, there are four logically possible configurations: dental-dental (t…t), dental-retroflex (t…ʈ), retroflex-dental (ʈ…t), and retroflex-retroflex (ʈ…ʈ). Let us assume that dental stops are more frequent overall than retroflex stops in the language. Moreover, let us assume that the language has a total of 500 roots containing non-adjacent C_1-C_2 pairs in which both consonants are coronal stops of one kind or another, and that the observed counts for each possible C_1-C_2 configuration are as shown in Table 5.

Table 5 Language X: Observed values (n = 500)

C_1\ C_2	t	ʈ
t	320	80
ʈ	80	20

The leftmost column in Table 5 represents the first member of the C_1-C_2 configuration, while the top row represents the second member of the configuration. Thus, t…t occurs 320 times in the lexicon, t…ʈ and ʈ…t occur 80 times each, and ʈ…ʈ occurs just 20 times. Based on the observed counts, we might be tempted to think that Language X has a co-occurrence restriction in which dental-dental configurations are highly favoured, while retroflex-retroflex configurations are strongly avoided. However, this is an artifact of the overall frequency of dentals relative to retroflexes. This becomes evident when the statistical method described above is applied to the data in Table 5. Expected values and O/E ratios for each C_1-C_2 configuration are listed in Table 6. The O/E ratios reveal that Language X places no restriction on the co-occurrence of coronal consonants. Each configuration occurs exactly as expected under random co-occurrence of coronal consonants (O/E = 1.00), taking into account the relative frequency of each consonant in each position.

Table 6 Language X: Observed (O), Expected (E) and O/E values (n = 500)

C_1\ C_2		t	ʈ
	O	320	80
t	E	320	80
	O/E	1.00	1.00
	O	80	20
ʈ	E	80	20
	O/E	1.00	1.00

Now consider another hypothetical language, Language Y, which is exactly like Language X in every way, except with respect to observed values for C_1-C_2 configurations. Observed, expected and O/E values for Language Y are shown in Table 7.

Table 7 Language Y: Observed (O), Expected (E) and O/E values (n = 500)

C_1\ C_2		t	ʈ
	O	400	40
t	E	352	88
	O/E	1.14	0.45
	O	0	60
ʈ	E	48	12
	O/E	0.00	5.00

For Language Y, the statistics reveal an assimilatory co-occurrence restriction. Once again, dentals are more frequent than retroflexes overall. Nevertheless, dental-dental and retroflex-retroflex configurations, which agree in place of articulation, both occur more frequently than expected. In fact, retroflex-retroflex configurations occur five times more often than expected (O/E = 5.00). At the same time, dental-retroflex and retroflex-dental pairs, which disagree in place of articulation, occur less frequently than expected. Avoidance is categorical in the case of retroflex-dental configurations (O/E = 0.00) and gradient in the case of dental-retroflex configurations (O/E = 0.45). Thus, the statistical method is useful for highlighting both categorical and gradient co-occurrence restrictions.

There is an important limitation to the statistical method as applied to coronal co-occurrence restrictions in the context of South Asia. While the method is useful for demonstrating the existence of retroflex consonant harmony in languages that have it, it is not very useful for demonstrating the absence of retroflex consonant harmony in languages that lack it. This is because O/E ratios compare observed counts to values expected under random co-occurrence of segments. However, in the absence of retroflex harmony, the distribution of retroflex segments is rarely (if ever) random. Rather, retroflex consonants are typically subject to other independent phonotactic restrictions (see §2.3). For instance, many South Asian languages that lack retroflex consonant harmony prohibit retroflex segments in word-initial position. The near-categorical absence of word-initial retroflex segments in these languages leads to extremely low expected frequencies for configurations involving initial retroflexes (e.g., retroflex-dental and retroflex-retroflex). Low expected values, in turn, yield highly exaggerated and unreliable O/E ratios for those configurations.

This point can be illustrated with reference to yet another hypothetical language, Language Z, which is like Languages X and Y above, except that it avoids word-initial retroflexes. Observed, expected and O/E values for Language Z are shown in Table 8.

Table 8 Language Z: Observed, Expected and O/E values ($n = 500$)

$C_1 \backslash C_2$		t	ʈ
	O	400	99
t	E	399.2	99.8
	O/E	1.00	0.99
	O	0	1
ʈ	E	0.8	0.2
	O/E	0.00	5.00

Looking only at O/E ratios in Table 8, we might conclude that retroflex-retroflex configurations are highly favoured in Language Z because they occur five times more frequently than expected. This is misleading. There is, in fact, only one example. The exaggerated O/E value is an artifact of the extremely low expected value for configurations involving initial retroflexes. In this case, the expected value for retroflex-retroflex configurations is just 0.2. With an expected value this low, a single occurrence leads to an O/E ratio of 5.00. Although the data in Table 8 reflects a hypothetical language, it is representative of attested data patterns in many South Asian languages. As this example serves to show, the statistical method is not only unenlightening when applied to these languages, but also misleading. For this reason, O/E ratios are not computed for languages of this type. Instead, simple observed values are left to speak for themselves.

This leads to another important point about the methodology employed here: the most reliable indicator of retroflex consonant harmony is not necessarily an over-attested O/E ratio for retroflex-retroflex configurations. Nor is it an under-attested O/E ratio for retroflex-dental configurations, since these are often avoided for independent reasons in languages that lack retroflex consonant harmony. Rather, a more reliable indicator of harmony is an over-attested

O/E ratio for retroflex-retroflex configurations combined with a highly under-attested O/E ratio for disharmonic dental-retroflex configurations. This is evident when comparing Table 7 with Table 8. Both scenarios have an O/E value of 5.00 for retroflex-retroflex configurations, and an O/E value of 0.00 for retroflex-dental configurations. However, in Table 7, which represents a language with a gradient form retroflex consonant harmony, this is coupled with a relatively low O/E ratio for dental-retroflex configurations (O/E = 0.45). In Table 8, which represents a language with a restriction on initial retroflexion instead of harmony, dental-retroflex configurations occur more-or-less as expected (O/E = 0.99). Thus, the tell-tale sign of retroflex consonant harmony in the South Asian context is the favouring of retroflex-retroflex configurations combined with the avoidance of dental-retroflex configurations.

In sum, statistical methods are useful for highlighting long-distance co-occurrence restrictions on consonants that might otherwise go unnoticed. In subsequent chapters, they are applied to the study of coronal co-occurrence restrictions in South Asian languages, bearing in mind the limitations mentioned above.

1.5 Overview of the dissertation

This chapter has laid the groundwork for the study of retroflex consonant harmony in South Asia by addressing foundational questions such as: What is retroflexion? What is consonant harmony? And what are the languages of South Asia? In addition, it has introduced the statistical method used to explore long-distance co-occurrence restrictions on consonants in subsequent chapters. The remainder of the dissertation is structured as follows.

Chapter 2 sets the stage for the study of retroflex consonant harmony in South Asia by establishing important facts concerning retroflexion in South Asia. Among these are the fact that retroflexion is a prominent areal feature of the region, one that occurs frequently in languages of all South Asian families. Another is that retroflexion is often subject to phonotactic restrictions. The most common restriction is a prohibition on word-initial retroflexes, which is characteristic of many Dravidian and Indo-Aryan languages, and possibly some Munda languages. A claim made in Chapter 2 is that phonotactic restrictions on retroflexion are a direct result of the evolution of retroflexion in a language. The phonotactic restrictions on retroflexion, and their historical basis, are both relevant to the study of retroflex consonant harmony because harmony has the effect of introducing a large number of word-initial retroflexes in languages that lacked them historically. In most cases, consonant harmony is not the only source of word-initial retroflexion, but it is clearly a major contributing factor.

The main empirical thrust of the dissertation is Chapter 3, which presents a survey of retroflex consonant harmony in South Asia. A major finding of the survey is that retroflex consonant harmony is a widespread areal trait affecting most languages in the northern half of the South Asian subcontinent, including languages from at least three of the four major South Asian families: Dravidian, Indo-Aryan and Munda. Retroflex consonant harmony in South Asia is manifested primarily in the form of static co-occurrence restrictions on roots (MSCs), which are the product of diachronic sound changes. Case studies are reviewed from a broad sample of languages. The cases of consonant harmony documented in Chapter 3 include some that have not been previously reported and others that have been reported but not adequately described. Most notably, the chapter documents retroflex consonant harmony in Indo-Aryan languages of the Dardic group, including Kalasha and Indus Kohistani, which have rich and typologically rare coronal consonant inventories. These languages exhibit striking similarity effects that have not been clearly observed in other retroflex consonant harmony systems. They figure prominently in all remaining chapters, where they have much to contribute to the discussion of theoretical issues surrounding consonant harmony.

The remaining chapters examine the theoretical issues introduced in §1.3.4 in light of the evidence from retroflex consonant harmony systems in South Asia. Chapter 4 addresses the theoretical debate concerning the mechanism(s) of assimilation. There, it is argued that the evidence from South Asia is largely consistent with the typological distinction between feature spreading and feature agreement, and provides support for the hypothesis that consonant harmony is the product of agreement, not spreading. Following through on this conclusion, Chapter 5 sketches a formal account of retroflex consonant harmony within the Agreement by Correspondence (ABC) framework of Rose & Walker (2004).

Chapter 6 discusses the evaluation of similarity in consonant harmony systems. While it is widely recognized that the set of interacting segments in consonant harmony systems is often constrained by similarity, the means by which languages evaluate similarity remains unclear. Chapter 6 reviews three different proposals and demonstrates that each one encounters problems when applied to retroflex consonant harmony systems in South Asian languages. The chapter makes no attempt to offer a new or definitive solution to the problem. Rather, it limits itself to identifying the challenges encountered by existing proposals and, wherever possible, suggesting possible ways forward.

Finally, Chapter 7 summarizes the main conclusions and contributions of the study and identifies areas for future research.

Chapter 2
Retroflexion in South Asia

South Asia has been identified as a 'linguistic area' or 'Sprachbund', i.e., a geographic area in which languages of different genetic stock have come to resemble each other through a history of contact and convergence. From a phonological point of view, the most prominent and widespread areal trait of South Asia is contrastive retroflexion. Most South Asian languages exhibit contrastive retroflexion regardless of their genetic affiliation (Emeneau, 1956; Ramanujan & Masica, 1969; Bhat D. N., 1973; Masica, 1992).

This chapter provides an overview of retroflexion in South Asia as a backdrop to the survey of retroflex consonant harmony in Chapter 3, which is the main empirical focus of the dissertation. §2.1 looks at retroflexion in South Asia from a synchronic perspective, addressing questions such as: How common is contrastive retroflexion in each South Asian language family (§2.1.1)? What coronal places of articulation are retroflex phonemes distinguished from (§2.1.2)? And what kinds of retroflex phonemes occur (in terms of manner), and how frequent is each type (§2.1.3)? §2.2 examines the ultimate origin and subsequent diachronic development of retroflexion in each South Asian family. §2.3 describes two important phonotactic constraints on retroflex segments and demonstrates that they arise more or less directly from the diachronic origins of retroflexion examined in §2.2. Finally, some points regarding the phonological behaviour of retroflex segments in local coronal assimilation are reviewed in §2.4.

2.1 Synchronic perspectives on retroflexion in South Asia

2.1.1 Retroflexion and language family

Retroflexion, as a contrastive property, is relatively rare cross-linguistically. For example, only about 11% of the world's languages have a distinctive series of retroflex stops (Maddieson, 1984, p. 32; Hamann, 2003, p. 3). Within South Asia, however, retroflexion is a prominent areal feature affecting the vast majority of languages (Emeneau, 1956; Ramanujan & Masica, 1969; Bhat D. N., 1973). Contrastive retroflexion occurs in each of the four major South Asian language families and also in some of the minor families and isolates. Figure 7 shows the extent of contrastive retroflexion in each family.[1]

[1] The statistics in Figure 7, and other statistics cited throughout this section, are based on an independent survey of phonological descriptions of 196 South Asian languages. The sample includes 35 Dravidian languages, 63 Indo-Iranian languages (61 Indo-Aryan, 2 Iranian), 18 Austro-Asiatic languages (15 Munda, 3 Mon-Khmer), 72 Tibeto-Burman languages, and 8 "Other" languages (3 Andamanese, 3 Tai-Kadai, 2 Isolates). The languages included in the survey and the data sources consulted for each language are listed in Appendix A.

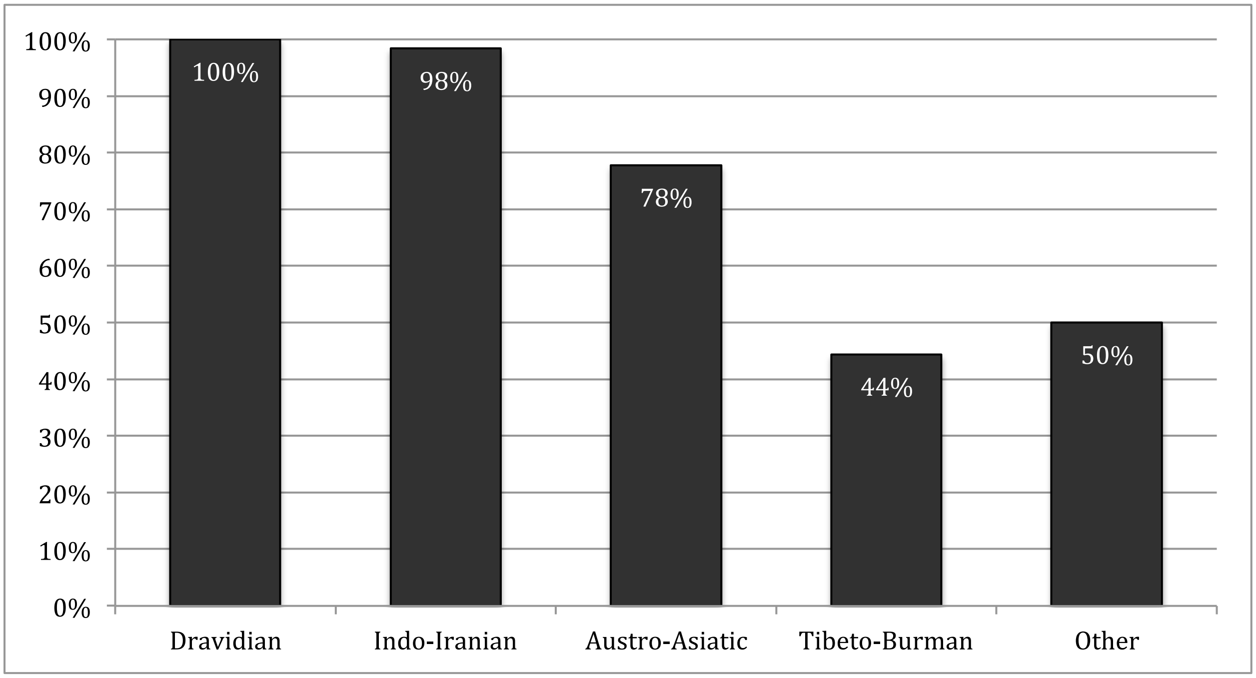


Figure 7 Percentage of languages with contrastive retroflexion in each South Asian language family

Contrastive retroflexion occurs in about 75% of all South Asian languages. The breakdown in Figure 7 demonstrates that retroflexion is most characteristic of Dravidian (100%), Indo-Iranian (98%) and Austro-Asiatic (78%). Within the Austro-Asiatic family, it occurs only in the Munda branch (93%), not in any of the Mon-Khmer languages. Retroflexion occurs least frequently in Tibeto-Burman (44%). Within the Tibeto-Burman family, retroflexion is found mostly in the Western Himalayish, Tamangic and Tibetan groups (all of which are subsumed under the larger Tibeto-Kanauri group), and only rarely in the other subgroups, most of which are concentrated in eastern Nepal and northeast India. Among the minor language families and isolates (subsumed under the label “Other” in Figure 7), retroflexion occurs only in the Andamanese languages and Burushaski.

In geographic terms, retroflexion is most prominent in the south, west and northwest, where the bulk of Dravidian and Indo-Iranian languages are spoken. It is least prominent in the northeast, where it is absent from most Tibeto-Burman languages, Khasi (Austro-Asiatic, Mon-Khmer), the Tai-Kadai languages of Assam, and even the Indo-Aryan language, Asamiya (a.k.a. Assamese).

In sum, contrastive retroflexion is a prominent areal feature of South Asia. It occurs in all Dravidian languages, the vast majority of Indo-Iranian (Indo-Aryan) and Munda languages, and a minority of Tibeto-Burman languages. The following sections look at retroflexion as it relates to other coronal places of articulation (§2.1.2) and as it relates to manner of articulation (§2.1.3) within these language families.

2.1.2 Retroflexion and coronal place systems

South Asian languages distinguish anywhere from three to six places of articulation. Like most languages of the world, they tend to distinguish more places of articulation within the class of stops (i.e., non-continuant obstruents, whether plosive or affricate) than in any other manner class. For this reason, the full range of place contrasts is best evaluated by examining the class of stops in each language. Figure 8 shows the overall frequency of stops at each place of articulation in South Asian languages. [2]

---

[2] The focus here is on contrast, not on phonetic realization. For this reason, dental and alveolar stops are grouped together in the "Dental or Den(ti)-Alv(eolar)" class in Figure 8 except in those cases where they contrast. Thus, the "Alveolar" category includes only those cases where apico-alveolar stops contrast with dentals, while the "Dental or Den(ti)-Alv(eolar)" category includes all those stops described as such, plus a few that are described as non-distinctively alveolar (e.g., Asamiya and some TB languages).

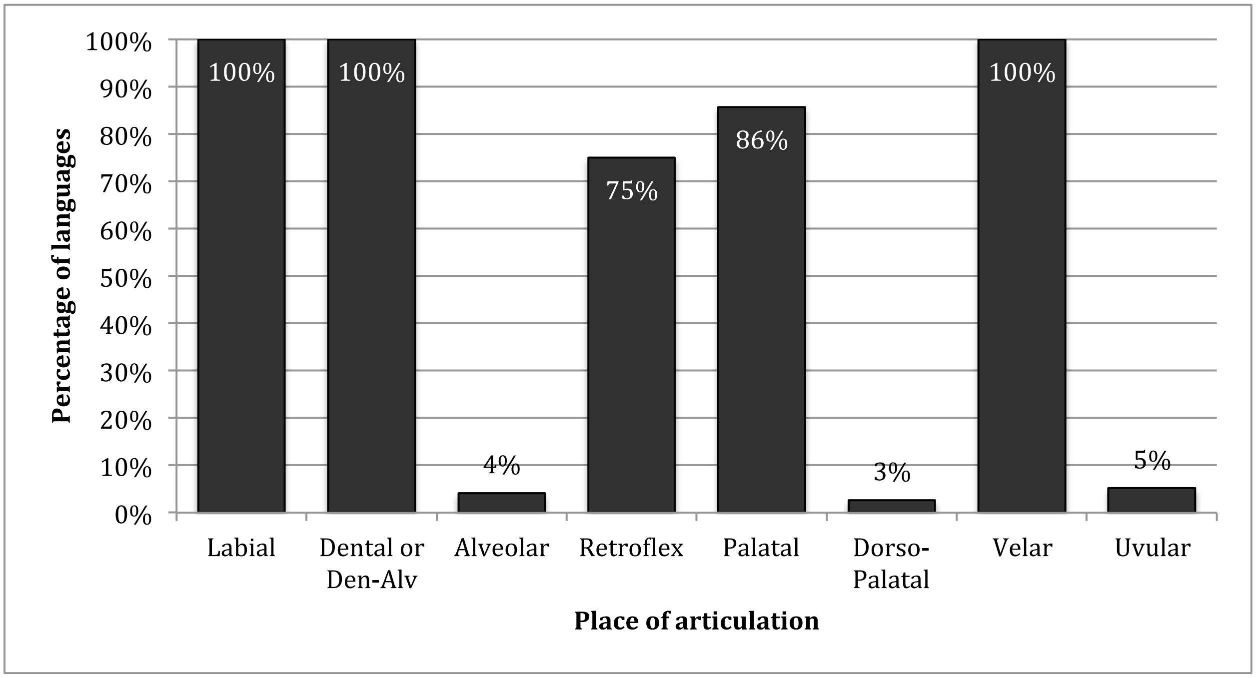


Figure 8 Percentage of South Asian languages contrasting stops at each place of articulation

Minimally, all South Asian languages distinguish labial and velar consonants from a coronal denti-alveolar series (100%). The vast majority of them also distinguish a retroflex (75%) and palatal series (86%). As a result, the statistically dominant pattern is one with five distinct places of articulation, three of which are coronal (Ramanujan & Masica, 1969, p. 564). The dominant three-way coronal place system includes dental, retroflex and palatal places, as shown in (1).

(1) Statistically dominant 3-way coronal place contrast in South Asian languages

Dental	Retroflex	Palatal
t	ʈ	ʧ

In three-way coronal place systems like that in (1), the dental series is typically lamino-dental for plosives but may be more apico-alveolar for sonorants. In such cases the series can be described broadly as denti-alveolar since no contrast exists between dental and alveolar places and the specific place realization of each member in the series is dependent on its manner. The palatal series is typically laminal post-alveolar and invariably realized with some affrication. In some languages it may tend toward a more lamino-alveolar realization or may vary between post-alveolar and alveolar realizations depending on vocalic context. Where conditioned variation occurs, the post-alveolar articulation typically occurs before front vowels and the alveolar articulation elsewhere.[3] Similarly, the retroflex series may vary somewhat depending on language family and/or vocalic context. The retroflex consonants of Indo-Aryan languages are often described as less retroflex than those of Dravidian (e.g., Ladefoged & Bhaskararao, 1983; Ladefoged & Maddieson, 1996). However, retroflex consonants in languages of both families may vary depending on speaker, speech rate, manner of articulation, or vocalic context. Retroflex consonants tend to be more apico-alveolar in the context of front vowels and more sub-apical post-alveolar in the context of back vowels (Reddy K. N., 1986; Dixit & Flege, 1991; Dart & Nihalani, 1999; Khatiwada, 2007).

The three-way coronal place system in (1) is reported in approximately 86% of all Dravidian, Indo-Iranian and Austro-Asiatic languages (combined), and in about 31% of all other South Asian languages (i.e., Tibeto-Burman and the minor families and isolates

---

[3] The complementary distribution of alveolar and palato-alveolar affricates is an areal trait affecting a broad belt of languages stretching from coast to coast across south-central India, roughly between 15^0 and 20^0 latitude (Sreekantaiya, 1954; Emeneau, 1956; Ramanujan & Masica, 1969). A similar pattern also occurs in many Tibeto-Burman languages of Nepal.

combined).[4] All cases of retroflex consonant harmony discussed in Chapter 3 occur in languages with coronal place systems of this kind.

A few South Asian languages have smaller coronal place systems that lack retroflex consonants (e.g., Sora (Munda)), palatal consonants (e.g., Konda (Dravidian)),[5] or both (e.g., Bodo (Tibeto-Burman) and Asamiya (Indo-Aryan)). A few Dravidian languages maintain a maximal four-way coronal place system that includes a distinctive apico-alveolar series in addition to dental, retroflex and palatal, as shown in (2).

(2) Maximal 4-way coronal place contrast in some South Dravidian languages

Dental	Alveolar	Retroflex	Palatal
t̪	t̺	ʈ	ʧ

The four-way coronal place system in (2) has been reconstructed for Proto-Dravidian (Steever, 1998b; Krishnamurti, 2003) and is reported in about 23% of Dravidian languages, all

---

[4] Some Tibeto-Burman languages with two-way coronal place systems distinguish plosives and affricates at the denti-alveolar place. Although the affricates are described as predominantly denti-alveolar they are often reported to have palato-alveolar allophones, or to differ from the plosives along the apical/laminal dimension. Given the allophonic variation and the potential difference in active articulators, it might be possible to treat these affricates as constituting a distinct 'place' series. If this is so, then many Tibeto-Burman languages classified here as having two coronal places would qualify as having three, thus bringing the Tibeto-Burman family somewhat more in line with the dominant regional trends. However, even under this analysis, Tibeto-Burman family would still exhibit fewer three-way coronal place systems relative to the other families.

[5] In Konda, the palatal stops of Proto-Dravidian have developed into denti-alveolar fricatives and, thus, no longer constitute a unique place series.

of them belonging to the South Dravidian group.[6] Languages with this system constitute only about 4% of all South Asian languages (as shown in Figure 8 above). No cases of consonant harmony are known to occur in any of these languages.

In sum, most South Asian languages distinguish three coronal places of articulation: dental, retroflex and palatal. A few languages, most of them Tibeto-Burman, exhibit smaller coronal place systems that lack the retroflex series, or in some cases the palatal series, or both. A few South Dravidian languages maintain maximal four-way coronal place systems that include a distinctive apico-alveolar series in addition to the lamino-dental, retroflex and palatal series of most other South Asian languages.

2.1.3 Retroflexion and manner of articulation

All South Asian languages with contrastive retroflexion have retroflex plosives. Most of them extend the retroflex contrast to at least one other manner of articulation. Figure 9 shows the frequency of contrastive retroflexion in each manner class.

[6] South Dravidian languages reported to distinguish apical alveolars from both dentals and retroflexes include the following: old literary Tamil, the Kanniyakumari dialect of modern colloquial Tamil, Malayalam, Irula, Kota, Toda, Paniya and Urali.

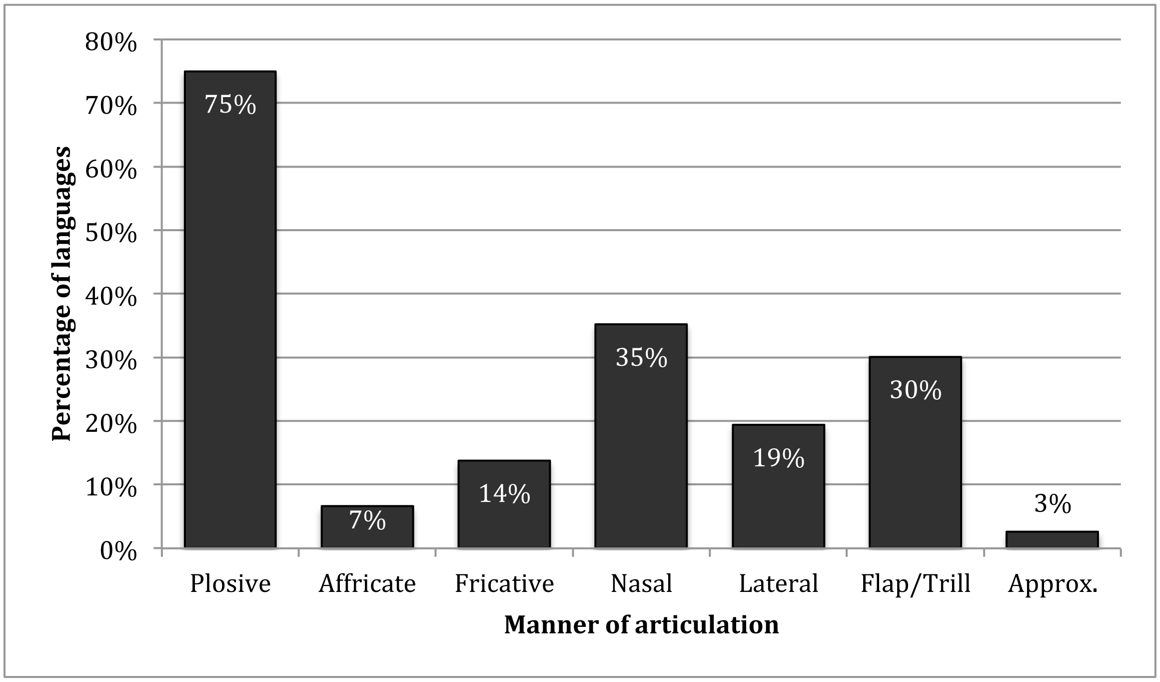


Figure 9 Contrastive retroflexion by manner class in South Asian languages

The statistics in Figure 9 reveal that retroflex plosives are by far the most common type of retroflex consonant. As a general rule, South Asian languages do not distinguish retroflex consonants in any other manner class without also distinguishing them in the class of plosives. After plosives, the next most frequent retroflex phonemes are the sonorants: nasals (/ɳ/), flaps (/ɽ/) and laterals (/ɭ/). Retroflex nasals occur phonetically in homorganic nasal + stop sequences in virtually all languages that have retroflex stops. However, their phonemic status is limited to those languages where they also occur independently in other environments. Similarly, retroflex flaps occur phonetically as lenis allophones of retroflex plosives (particularly voiced /ɖ/) in many languages. However, they have developed independent phonemic status in some cases. In broad geographic terms, languages with /ɽ/ and languages with /ɭ/ are nearly complementary. The phoneme /ɽ/ occurs more frequently in northern and central parts of the subcontinent while the phoneme /ɭ/ occurs primarily in southern areas.

On the whole, South Asian languages have very impoverished coronal sibilant systems. Thus, retroflex fricatives (/ʂ/), and affricates (/ʈʂ/) are relatively rare. They occur primarily in Indo-Aryan languages of the Dardic group in northern Pakistan and other languages of the northwest (e.g., Burushaski). Retroflex approximants (/ɻ/) are exceedingly rare but are reported, for instance, in Tamil, Malayalam and Burushaski.

At the very least, all of these retroflex phonemes stand in opposition to a non-retroflex denti-alveolar counterpart. In some cases they are further distinguished from palatal counterparts. Retroflex approximants are distinguished only from palatal (and labial) approximants because denti-alveolar approximants do not occur. However, if they are regarded as a kind of rhotic, as in Ladefoged & Maddieson (1996), then the retroflex approximants also stand in opposition to denti-alveolar /r/, which is typically a flapped or trilled alveolar rhotic in the South Asian context.[7]

Although they do not belong to the coronal consonant systems, retroflex vowels deserve some mention here. At least two South Asian languages have been reported to distinguish retroflex vowels: Badaga, a South Dravidian language spoken in the Nilgiri hills of southern India (Emeneau, 1939), and Kalasha, an Indo-Aryan language of the Dardic group in northern Pakistan (Mørch & Heegård, 1997; Heegård & Mørch, 2004). In both cases, the retroflex vowels have developed under the conditioning influence of adjacent retroflex consonants that were subsequently lost through lenition (or possibly merged with the vowels). However, the

[7] In some Tibeto-Burman languages the phoneme /r/ is realized as an approximant with retroflex qualities (much like English). However, in these cases the retroflexion is not necessarily a contrastive property since there is no non-retroflexed rhotic in the system.

nature and status of the contrast remains unclear in Badaga. Other Dravidian languages of the Nilgiri area are reported to have 'centralized' or 'back unrounded' vowels that are identified with the retroflex vowels of Badaga and stem from the same historical antecedents. Since different labels are used interchangeably for these vowels, it is not clear whether they qualify as 'retroflex' or merely as 'centralized' or 'retracted' (Zvelebil, 1973, p. 11; Diffloth, 1975, pp. 55, footnote 5). Moreover, the contrast appears to be lost altogether in most (or all) contemporary dialects of Badaga (Hockings & Pilot-Raichoor, 1992; Ladefoged & Maddieson, 1996, pp. 313–14; Krishnamurti, 2003, p. 51). The retroflex vowels of Kalasha and their relevance to the study of consonant harmony are discussed further in Chapter 3 (§3.3.2.3).[8]

In sum, contrastive retroflexion is an areal feature of South Asia affecting an overall majority of languages, most notably those of the Dravidian, Indo-Aryan and Munda families, but also, to a lesser extent, those of the Tibeto-Burman family. Most languages with contrastive retroflexion distinguish retroflex from dental (or denti-alveolar) and palatal coronals, though a few maintain smaller or larger coronal place systems. Contrastive retroflexion occurs most frequently within the class of plosives. Other common retroflex segments include nasals, flaps and laterals. Retroflex fricatives and affricates are relatively rare but do occur, particularly in languages of the northwest. Phonemic retroflex approximants and vowels are exceedingly rare.

---

[8] Qiang, a Tibeto-Burman language spoken in China, may also have phonemic retroflex vowels. In some cases, retroflexion may also spread from one vowel to another in vowel harmony (see LaPolla, 2003, p. 574).

2.2 Diachronic perspectives on retroflexion in South Asia

From a synchronic point of view, retroflexion is pervasive in South Asia. This section examines retroflexion in South Asia from a diachronic point of view and addresses the question of how this situation came to be. The most common cross-linguistic sources of retroflexion are introduced in §2.2.1, along with their phonetic grounding. The remainder of this section reviews the diachronic origins of retroflexion in each of the South Asian language families (§2.2.2 – 2.2.5). Phonotactic constraints on retroflex consonants and their relation to these diachronic developments are discussed in §2.3.

2.2.1 Cross-linguistic sources of retroflexion

Cross-linguistically, retroflex consonants typically emerge as phonetically conditioned variants of anterior coronal consonants. Various conditions are known to induce retroflexion. The most common of these are listed in (3) (Bhat D. N., 1973; Hamann, 2003; 2005).

(3) Common cross-linguistic sources of retroflexion

a.	retroflexion in liquid/rhotic context	e.g.,	rt / tr	>	ʈ
b.	retroflexion in back vowel context	e.g.,	ut / tu	>	uʈ / ʈu
c.	retroflexion of voiced (implosive) stops	e.g.,	d / ɗ	>	ɖ

Liquids are perhaps the single most frequent source of retroflexion cross-linguistically (3)(a). The liquids that induce retroflexion in adjacent denti-alveolar consonants are most often rhotics, though cases involving laterals are also well attested. This trend may be grounded in both articulatory and perceptual factors. In articulatory terms, rhotics are often apical and prone to developing postalveolar retroflex allophones, particularly when they are realized as approximants (Maddieson, 1984, p. 82; Hall, 1997a, pp. 215, footnote 21; Hamann, 2003, p.

87ff). Thus, they can induce an apical postalveolar articulation in neighbouring coronals through local assimilation. In acoustic terms, rhotics are often characterized by a low F3 that is reminiscent of retroflexion. Hamann (2003; 2005) argues that retroflexion in rhotic contexts can be perceptually motivated if the cues stemming from a rhotic are misparsed and attributed to a neighbouring coronal consonant, which is then perceived as retroflex. It is not immediately clear whether similar articulatory and acoustic trends hold for apico-alveolar laterals cross-linguistically. Whatever the case may be, variation between alveolar and retroflex laterals is well attested in South Asia, both synchronically and diachronically. Thus, their ability to induce retroflexion is not surprising in the South Asian context.

Liquids can induce retroflexion in preceding or following consonants. However, retroflexion occurs more often in post-liquid environments (i.e., rt / lt > ʈ) than in pre-liquid environments (i.e., tr / tl > ʈ). Hamann (2003) attributes this to the fact that retroflex place cues are more salient in VC transitions than CV transitions (see §1.2.3). Thus, the acoustic cues of a rhotic in the post-vocalic VrC context are easily misparsed as retroflexion of the following C. A similar development is possible but less likely in pre-vocalic CrV contexts where retroflex cues are less salient.

Back vowels are also known to induce retroflexion in adjacent denti-alveolar consonants (3)(b). This is especially true of rounded back vowels, such as /u/ and /o/, but also of unrounded back vowels such as /a/ in some cases. Once again, these trends are grounded in both articulatory and acoustic factors. In articulatory terms, back vowels can induce retraction of anterior coronals through co-articulation. In acoustic terms, the tongue retraction and lip rounding associated with back vowels can contribute to a lowering of F3, which can be reconstrued as retroflexion on an adjacent coronal consonant (Hamann, 2003).

Bhat (1973) identifies implosion as another articulatory property that can induce retroflexion in denti-alveolar stops. However, Hamann & Fuchs (2010) have argued that this trend is not necessarily tied to airstream mechanism, but to voicing (3)(c). They demonstrate that the articulatory and aerodynamic requirements for voiced alveolar or dental stops, whether plosive or implosive, can lead to tongue tip retraction and tongue mid lowering, thereby causing retroflexion in voiced coronal stops.

Although it is not discussed in any of the sources consulted, the phonetic tendency described by Hamann & Fuchs (2010) can be regarded as part of a larger trend in which coronal place of articulation is often conditioned to some degree by manner of articulation. More specifically, coronal place is often correlated with sonority. The less sonorant the manner, the more laminal and anterior it is likely to be; the more sonorant the manner, the more apical and posterior it is likely to be. As noted earlier (§2.1.2), denti-alveolar plosives are overwhelmingly lamino-dental in South Asian languages, whereas nasals and liquids tend to be apico-alveolar and approximant rhotics are often retroflex. If voiced plosives are regarded as more sonorant than their voiceless counterparts, then the tendency for voiced coronal plosives to be more apical and posterior relative to their voiceless counterparts can be viewed as another instance of this overall trend. The same trend may also be reflected in the tendency for liquids to induce retroflexion. Since liquids are among the most sonorous of oral consonants (Ladefoged & Maddieson, 1996, p. 182), they are also the most inclined toward a postalveolar retroflex articulation.

In sum, retroflexion can emerge through the assimilatory influence of a liquid (rhotic or lateral) or back vowel, or through the articulatory and aerodynamic requirements of voicing in denti-alveolar plosives. The following subsections review the diachronic origins of retroflexion in each of the South Asian language families (§2.2.2 – 2.2.5). All of the developments identified above are attested within South Asia. However, it will be seen that the two largest families, Dravidian and Indo-Aryan, have developed retroflexion primarily in post-liquid (i.e., rt / lt > ʈ) and post-vocalic environments (i.e., ut > uʈ), while Tibeto-Burman languages have developed them primarily in pre-rhotic contexts (i.e., tr > ʈ). Munda languages show evidence of retroflexion induced by voicing of denti-alveolar plosives.

2.2.2 Dravidian

The Indo-Aryan, Munda and Tibeto-Burman families all have at least some member(s) within South Asia that lack retroflexion. Moreover, they all belong to larger families in which retroflexion is not a characteristic property outside of the South Asia region (i.e., Indo-European, Austro-Asiatic and Sino-Tibetan, respectively). In contrast to this, the Dravidian languages have no established relatives outside of South Asia and all Dravidian languages exhibit contrastive retroflexion without exception. As a result, retroflexion is regarded as a native property of Proto-Dravidian. The consonants of Proto-Dravidian are listed in (4).

(4) Proto-Dravidian consonant phonemes (Steever, 1998b, p. 14; Krishnamurti, 2003, p. 91)

LAB	DEN	ALV	RET	PAL	VEL	GLOT
p	t	t̠	ʈ	ʧ	k	
m	n	[n̠]	ɳ	ɲ	[ŋ]	
		l	ɭ			
		r				
w			ɻ	j		(h)

The phonological system posited for Proto-Dravidian in (4) is almost identical to that of Old Tamil, which is well supported by a long literary tradition, with written records dating back more than two thousand years (Lehmann, 1998). Proto-Dravidian is reconstructed with a maximal four-way coronal place system that included lamino-dental, apico-alveolar, retroflex and palatal stops. All oral stops were voiceless initially and in gemination, and voiced after nasals, which were always homorganic ([mb], [nd], [n̠d̠], [ɳɖ], [ɲʤ], [ŋg]). Intervocalic stops were both voiced and lenis. Lenition of apico-alveolar */t̠/ produced an alveolar trill, commonly transliterated [r̠] in the Dravidian tradition, which was distinct from the alveolar flap */r/. Proto-Dravidian */n/ was dental initially and before dental plosives and alveolar [n̠] elsewhere.

Since retroflexion is considered native to Dravidian, the existence of retroflex segments in Proto-Dravidian is generally assumed and the question of their ultimate origin within the family is rarely addressed.[9] There is, however, some historical-comparative evidence and typological evidence that bears on this issue. Although many details remain unclear, there is a

[9] For instance, Caldwell (1875) writes: “the Dravidian languages, which claim to have had an origin independent of Sanskrit, and which appear to have been spoken throughout India prior to the arrival of the Aryans, possess the lingual sounds in question [i.e., retroflex consonants], and, for aught that appears, were in possession of them always” (*Comparative grammar*, Part I, p. 32; cf., Caldwell, 1856, p. 112).

general consensus that retroflexion first emerged phonetically in the class of liquids and spread from liquids to following nasals and plosives through local assimilation across morpheme boundaries (Zvelebil, 1970, pp. 101–104, 174–175; Tikkanen, 1999; Levitt, 2010). The general hypothesis is sketched in (5).

(5) Sources of retroflexion in (pre-) Proto-Dravidian

a. retroflexion of liquids: **l > *l, ɭ, ɻ

b. coronal assimilation: *lt > *n̺t̺, *t̺(t̺)

*ɭt, ɻt > *ɳʈ, *ʈ(ʈ)

Levitt (2010, p. 63), citing previous work by Devaneyan (1966), argues that the three liquids of Proto-Dravidian, */l, ɭ, ɻ/, can be traced to an original **/l/ in the pre-Proto-Dravidian stage. This is sketched in (5)(a). The original **/l/ developed a retracted retroflex allophone [ɭ], which in turn developed an approximant allophone [ɻ]. Subsequent changes led to a phonemic split. Levitt does not elaborate on the conditions that produced the original allophonic variation or the developments that led to a phonemic split. A solution to this question is beyond the scope of the present study. It is worth noting, however, that developments of the kind in (5)(a) are attested elsewhere under various conditions. For instance, liquids are highly prone to retroflexion cross-linguistically, especially in the context of back vowels (Bhat D. N., 1973, pp. 48–50). Also, a phonemic split between denti-alveolar and retroflex laterals occurred in many New Indo-Aryan languages. In these cases, Middle Indo-Aryan intervocalic /-l-/ developed into retroflex /-ɭ-/, while MIA geminate /-ll-/ became the new denti-alveolar singleton /-l-/ (Masica, 1991, p. 193).

Within the South Dravidian group, there is widespread alternation between /ɭ ~ ʈ(ʈ) ~ ɳʈ/ and /ɻ ~ ɳʈ/, and also between /l ~ t̺(t̺) ~ n̺t̺/. In these cases, all sources agree that the liquids /l, ɭ, ɻ/ are original, while stem forms containing nasals and plosives are the result of affixation and assimilation. This is sketched in (5)(b), where stem final liquids come into contact with suffixal */-t/ (or possibly */-nt/; Zvelebil, 1970, pp. 174–175). Under these conditions, alveolar */l/ yielded alveolar plosives and nasals (e.g., */l + t/ > */lt̺/ > */n̺t̺/ or */t̺t̺/ or */t̺/), whereas the retroflex liquids */ɭ, ɻ/ yielded retroflex plosives and nasals (e.g., */ɭ + t/ > */ɭʈ/ > */ɳʈ/ or */ʈʈ/ or */ʈ/). Morphophonological alternations of these (and other) types are attested in old literary Tamil (see Levitt, 2010, pp. 64–69 for examples). They are also reflected in stem alternations in contemporary South Dravidian languages (see Zvelebil, 1970, pp. 101–104 and 174–175 for examples). The loss of the conditioning alveolar or retroflex liquid, coupled with the preservation of an original dental */t/ in non-liquid environments, ultimately produced the three-way contrast between dental, alveolar and retroflex plosives and nasals found in Proto-Dravidian and Old Tamil.

Cross-linguistic typological evidence supports the hypothesis that apico-alveolar and retroflex phonemes can evolve from post-liquid dentals, as proposed for Proto-Dravidian. For instance, some Central Norwegian and Swedish dialects have developed a three-way contrast between dental, alveolar and retroflex plosives under conditions nearly identical to those proposed for Proto-Dravidian. In these dialects, dental /l/ developed a ‘dark’ retracted allophone [ɫ] in post-vocalic contexts. Retracted [ɫ] triggered retroflexion in following dental plosives (e.g., ɫt > ɫʈ > ʈ), while the rhotic /r/ produced alveolar plosives under the same conditions (e.g., rt > rt̺ > t̺). Once again, preservation of the original dental plosive in non-liquid environments, combined with the loss of the liquids that conditioned alveolar and retroflex plosives, resulted in a three-way contrast between dental, alveolar and retroflex plosives. These developments are summarized by Hock (1991) as shown in (6).[10]

(6) Development of coronal contrast in Norwegian & Swedish dialects (Hock, 1991, p. 79)

a.	Starting point:	t, d	rt, rd	łt, łd
b.	Assimilation:	---	rt̺, rd̺	łʈ, łɖ
c.	Loss of conditioning environment:	t, d	t̺, d̺	ʈ, ɖ

It is worth noting that the phonological systems of Proto-Dravidian and Old Tamil bear a striking resemblance to those of Australian languages, which also tend to distinguish dental, alveolar and retroflex stops and nasals.[11] Given the strong typological relation between Dravidian and Australian, it is worth considering whether their phonological systems developed along similar lines. Proto-Australian is generally reconstructed with two coronal series, one

---

[10] Apart from identifying them as "central" dialects, Hock (1991) does not provide details about the Norwegian and Swedish dialects that he has in mind, nor does he cite any sources. Most other accounts of coronal contrast in Norwegian and Swedish report a two-way distinction between dental and retroflex, not the three-way distinction involving dental, alveolar and retroflex reported by Hock. In all of these other accounts, the retroflex consonants derive historically from /r/ + dental sequences (Kristoffersen, 2000; Hamann, 2003; 2005; Simonsen, Moen, & Cowen, 2008).

[11] Evans (1995) identifies five phonological traits that are characteristic of Australian languages, all of which are shared by Proto-Dravidian and Old Tamil. They are: (i) six series of stops, each with a corresponding nasal, including labial, dental, alveolar, retroflex, palatal and velar (though not all nasals are phonemic in Proto-Dravidian and Old Tamil); (ii) a lack of laryngeal contrasts; (iii) a complete lack of phonemic fricatives; (iv) a distinction between two rhotics (i.e., alveolar /r/ and a retroflex approximant /ɻ/); and (v) phonemic vowel length. Many of these shared traits are rare cross-linguistically. This is especially true of the three-way coronal place distinction between dental, alveolar and retroflex consonants, the presence of two rhotics, and the complete absence of fricatives. Languages of the two families also exhibit similar phonotactic patterns. For instance, both avoid word-initial apical consonants (see discussion in §2.3). Phonological parallels between the Dravidian and Australian families have led some to contemplate a possible genetic relationship between the two (e.g., see Caldwell, 1856, pp. 51-53; Dixon, 1980, pp. 236-37; Levitt, 2010).

laminal and one apical. Thus, the origin of the apical vs. laminal contrast is not generally addressed for Australian. However, it is widely recognized that the retroflex series developed from the apico-alveolar series after rhotics, though back vowels may also have played a role (Dixon, 1980; 2002). Once again, there is support from typologically related languages for the development of retroflex segments from non-retroflex coronals in post-liquid environments.

The apico-alveolar series of Proto-Dravidian has been lost in all but a few conservative South Dravidian languages. Languages that preserve the apico-alveolar stops (in addition to the other coronal stops) include Malayalam, Irula, Kota, Toda, Paniya, Urali and the Kanniyakumari dialect of modern colloquial Tamil. In Roman transcriptions of these languages it is often represented as /r̠/ to reflect its lenis intervocalic realization, even though it is still realized as a plosive in gemination (i.e., /r̠r̠/ → [t̺t̺]) and post-nasally in most cases (i.e., /nr̠/ → [n̺d̺]). Outside of this small conservative group the alveolar stop has merged with other phonemes – primarily with /r/ intervocalically and with dental or retroflex plosives in other environments. As a result, most contemporary Dravidian languages exhibit the statistically dominant three-way coronal place system that includes dental, retroflex and palatal stops, without a distinctive alveolar series.

Proto-Dravidian */ɳ/ and */ɭ/ are preserved in most contemporary South Dravidian languages. Elsewhere they have tended to merge with their denti-alveolar counterparts, /n/ and /l/. In some cases, particularly in the South-Central group, */ɭ/ has developed into a non-lateral retroflex flap /ɽ/. The retroflex approximant */ɻ/ is preserved only in dialects of Tamil and Malayalam (both South Dravidian). Common reflexes in other Dravidian languages include /r/, /ɭ/, /ɽ/, /j/, /ɖ/ or ∅ (i.e., complete elision).

In sum, retroflexion is a native feature of Dravidian that can be traced back to Proto-Dravidian. Proto-Dravidian distinguished four coronal places of articulation including lamino-dental, apico-alveolar, retroflex and palatal. Although the ultimate origins of these contrasts are uncertain, evidence suggests that retroflexion developed first in liquids, and subsequently spread via local assimilation to stops and nasals in post-liquid environments. The apico-alveolar series has been lost in all but a few South Dravidian languages. As a result, most contemporary Dravidian languages now distinguish only three coronal places of articulation: dental, retroflex and palatal.

2.2.3 Indo-Aryan

Retroflexion was not a native feature of Proto-Indo-European. It was an innovation within the Indo-Aryan branch that developed only after the Aryans arrived in South Asia (c. 1500 BC). The arrival of the Aryans in South Asia brought them into contact with Dravidian populations who probably occupied much of the Indian subcontinent at that time. It is generally assumed that the development of retroflexion in Indo-Aryan was fostered by contact with Dravidian and supplemented by Dravidian loanwords. However, retroflexion in Old-Indo-Aryan can also be traced to internal developments stemming from Proto-Indo-European. To what extent these internal developments were influenced by Dravidian contact remains unclear. The main stages of internal development are summarized in (7) (Misra B. G., 1967, pp. 28–29, 63ff; Bhat D. N., 1973; Hamp, 1996; Tikkanen, 1999).

(7) Sources of retroflexion in Proto-Indo-Aryan

a. ruki: IE s, z > IA ʂ, ʐ / r, u, k, i ___

b. n-retroflexion: IA n > ɳ / r, ʂ (…) ___

c. coronal assimilation: IA ʂt, ʐd > ʂʈ, ʐɖ

d. sibilant laxing: IA ʐɖ > i^{r}ɖ, u^{r}ɖ > iɖ, uɖ

e. Fortunatov's law: IE lt, ls, ln > ɭt, ɭs, ɭn > ɭʈ, ɭʂ, ɭɳ > IA ʈ, ʂ, ɳ

The so-called 'ruki' rule in (7)(a) produced retroflex sibilants from IE *s and its voiced allophone *[z] when they occurred following a rhotic (i.e., /r/ or any of its allophones), a back vowel (i.e., /u/), a velar consonant (i.e., /k/ or any of its voiced or aspirated counterparts), or the vowel /i/.[12] The resulting retroflex sibilants, /ʂ/ and *[ʐ], together with Indo-Aryan /r/, subsequently spread retroflexion to any following dental nasal, whether adjacent or non-adjacent (7)(b) (cf. §3.2.1.1). The retroflex sibilants also spread retroflexion to a following dental stop via local coronal assimilation, thereby producing homorganic consonant clusters such as [ʂʈ] and *[ʐɖ] (7)(c). Subsequent developments led to the phonemicization of what were originally allophonic retroflex segments. For instance, the voiced retroflex sibilant that conditioned retroflexion in *[ʐɖ] clusters was lost to lenition, leaving the retroflex plosive /ɖ/ to

[12] The 'ruki' rule has been a perennial problem for phonological theories because it is not clear how the segments /r, u, k, i/ constitute a natural class or how they could all condition retroflexion. In particular, the inclusion of /i/ in the class is problematic because high front vowels are more commonly associated with de-retroflexion and palatalization of coronals, not with retroflexion. The most plausible explanation is that the ruki rule originally conditioned a range of allophones in PIE */s/. These may have included retroflex [ʂ] after /r, u/, palatal [ʃ] after /i/ and either retroflex [ʂ] or velar [x] after /k/. These allophones were later merged and phonemicized as retroflex [ʂ] in Indo-Aryan. Evidence for this hypothesis comes from the fact that the same 'ruki' environment has yielded different results in other branches of Indo-European; for instance, palato-alveolar [ʃ] in Iranian and Baltic and velar [x] in Slavic. For details and references see the discussion in Hamann (2003, pp. 107–111). For an alternative account see Hall (1997a).

be re-interpreted as a phoneme (7)(d). Similarly, historical-comparative evidence suggests that IE *l triggered retroflexion in following dental stops, sibilants and nasals (i.e., lt, ls, ln > ɭt, ɭs, ɭn > ɭʈ, ɭʂ, ɭɳ) before succumbing to elision (or merger with the following consonant) and leaving /ʈ/, /ʂ/, and /ɳ/ to stand as independent retroflex phonemes (7)(e).[13]

The net result of the developments in (7) was a phonological system with retroflex stops, fricatives and nasals, as attested in Old-Indo-Aryan Sanskrit.[14] The consonant phonemes of Sanskrit are shown in (8).

(8) Consonant phonemes of OIA Sanskrit (Whitney, 1993 [1889]; Cardona, 2003)

LAB	DEN	RET	PAL	VEL	GLOT
p	t	ʈ	ʧ	k	
pʰ	tʰ	ʈʰ	ʧʰ	kʰ	
b	d	ɖ	ʤ	g	
bʰ	dʰ	ɖʰ	ʤʰ	gʰ	
	s	ʂ	ʃ		(h) ɦ
m	n	ɳ	ɲ	ŋ	
	l				
	r				
v			j		

[13] Note the similarity between Fortunatov's law in Indo-Aryan, sketched in (7)(e), and the development of retroflexion in Dravidian, sketched in (5). Levitt (2010) argues that the application of Fortunatov's law in Indo-Aryan is the product of a Dravidian substratum, i.e., the result of Dravidian populations adopting Indo-Aryan speech and applying the phonetic and phonological patterns of their native Dravidian language to Indo-Aryan vocabulary.

[14] Early Vedic Sanskrit also distinguished the retroflex laterals [ɭ] and [ɭʰ], at least orthographically. However, these can be regarded as intervocalic allophones of /ɖ/ and /ɖʰ/ (Masica, 1991, p. 161).

Sanskrit had a rich system of coronal consonants that included stops, fricatives and nasals at three places of articulation: dental, retroflex and palatal. The rhotic /r/ is described variously as dental, alveolar or retroflex (Cardona, 2003, p. 109). Both /r/ and /l/ had syllabic counterparts, /r̩, l̩/. They are traditionally treated as part of the vowel system. The glottal continuant, commonly transliterated *h*, was in fact breathy voiced /ɦ/. The segment commonly transliterated *ḥ* corresponds to orthographic *visarga*, which represents a voiceless glottal fricative (IPA [h]) that can be treated as an allophone of /s/.

Retroflex plosives were relatively rare in early Vedic Sanskrit but became increasingly numerous in post-Vedic Sanskrit literature and in the Middle Indo-Aryan period. MIA developed geminate consonants from OIA C_1C_2 sequences, typically via regressive assimilation (e.g., OIA /sapta/ 'seven' > Pāli /satta/). In some cases this resulted in new retroflex plosives from dentals following /r/, as shown in (9).

(9) MIA retroflex plosives from OIA -rt, -rd (Masica, 1991, p. 176)

Skt. varti	>	Pkt.	vaʈʈi, vatti	'wick'
Skt. ard^{h}a	>	Pkt.	aɖɖ^{h}a, add^{h}a	'half'

Another source of new retroflex stops in MIA was the OIA sequence /st^h/, which often yielded retroflex plosives in both initial and non-initial positions, as shown in (10).

(10) MIA retroflex plosives from OIA st^h (Masica, 1991, pp. 172, 177)

Skt. st^{h}aːna	>	Pkt.	ʈ^{h}aːna	'place'
Skt. aːst^{h}aː	>	Pkt.	aʈʈ^{h}aː, att^{h}aː	'condition'

Further developments in the Middle and New Indo-Aryan periods either neutralized some retroflex contrasts or extended them in new ways. Several cases deserve mention. First of all, the dental laterals of MIA have developed into retroflex laterals intervocalically in some NIA languages (i.e., MIA /-l-/ > NIA /-ɭ-/). Meanwhile, their geminate counterparts have remained dental and have reduced to singletons (i.e., MIA /-ll-/ > NIA /-l-/). These developments have resulted in contrast between dental and retroflex laterals in some NIA languages, including Gujarati, Marathi, various 'Rajasthani' languages, and dialects of Panjabi, among others (Masica, 1991, p. 193).

Secondly, the voiced retroflex plosives /ɖ/ and /ɖʰ/ have been subject to lenition throughout NIA, most notably (but not exclusively) in intervocalic position. Lenition of these plosives has produced sonorant retroflex flaps, either as allophones of the plosives or (via subsequent developments) as independent retroflex phonemes, /ɽ/ and /ɽʰ/.

Thirdly, the three-way coronal contrast between dental, retroflex and palatal sibilants was neutralized to dental /s/ in most MIA dialects, but to palatal /ʃ/ in the Magadhi dialect, from which most NIA languages of the eastern zone have descended (Chatterji, 1970 [1926], pp. 92, 245; Masica, 1991, pp. 168, 186). These developments are still reflected in most NIA languages. However, the three-way contrast has been fully preserved in the Dardic languages of the Northwestern zone. These same languages have also developed a new series of retroflex affricates, /ʈʂ/ and /ʈʂʰ/, which derive primarily from OIA /kʂ/ sequences. Outside of the Dardic group, OIA /kʂ/ yielded aspirated velars or palatals in NIA (i.e., /kʰ/ or /ʧʰ/) (Masica, 1991, pp. 173, 177, 201). These developments are illustrated in (11), where Kalasha and Indus Kohistani (IK) represent the Dardic group and Hindi represents mainstream NIA.

(11) NIA retroflex affricates from OIA kʂ in Dardic (e.g., Kalasha and IK)

	OIA	MIA	Hindi	Kalasha	IK
‘milk’	kʂiːra	k^hiːra, $ʧ^h$iːra	k^hiːr	$ʈʂ^h$ir	$ʈʂ^h$íːr
‘wing, side’	pakʂa	pakk^{h}a	paːk^h	paʈʂ	pʌ̀ʈʂ^{hi}

Some Dardic languages, including Indus Kohistani and Shina, developed additional retroflex affricates and/or fricatives from C + r sequences in which C was a labial or dental consonant. Other Dardic languages, such as Kalasha and Palula, have generally retained OIA Cr sequences intact. Sindhi is unique among NIA languages in retroflexing dental plosives before /r/ without loss of the rhotic (i.e., tr > ʈr). Elsewhere, mainstream NIA has tended to simplify all Cr sequences through loss of the rhotic without retroflexion. These developments are illustrated in (12). Once again, Hindi represents mainstream NIA. Indus Kohistani (IK) represents those Dardic languages that developed retroflex affricates or fricatives from OIA Cr sequences, and Kalasha represents those Dardic languages that did not.

(12) NIA retroflex plosives (in Sindhi) and affricates (in IK) from OIA Cr-

	OIA	Hindi	Sindhi	Kalasha	IK
‘flea’	pluʂi, *priʂu	pissuː	—	priʂu	ʈʂìːʂ
‘brother’	b^hraːtr̩	b^haːiː	b^haːu, b^haːiː	baja	$ʐ^h$àː
‘three’	trajaḥ, triːɳi	tiːn	ʈre	tre	ʈʂàː
‘grape’	draːkʂaː	daːk^h	ɖraːk^{h}a	draʈʂ	ʐàːʈʂ

In sum, retroflexion has a long and complex history within Indo-Aryan. However, at the earliest stages, retroflexion emerged primarily in post-rhotic, post-vocalic and post-velar contexts (i.e., the ‘ruki’ rule), and also in post-lateral contexts (i.e., Fortunatov’s law).

Subsequent developments in OIA and MIA mostly involved the progressive spread of retroflexion from rhotics and sibilants to nasals and plosives. Later developments in some NIA languages involved the expansion of retroflex contrasts to new manners of articulation. In the case of the Dardic group, this included the development of retroflex affricates from OIA /kʂ/ sequences and, in some languages, from pre-rhotic stops (e.g., /tr/ > /ʈʂ/).

2.2.4 Munda

Little is known about the history of the Munda languages. Unlike Dravidian, Indo-Aryan and Tibeto-Burman, the historical study of Munda does not have the benefit of a long literary tradition. Most Munda languages remain largely unwritten, under-documented and endangered. Thus, reconstructions of Proto-Munda are somewhat tenuous compared to those of other South Asian families. Nevertheless, there is general consensus that contrastive retroflexion was not a feature of Proto-Munda.

The most widely accepted position on the Proto-Munda coronal system posits a single denti-alveolar series, possibly in opposition to a palatal series. The denti-alveolar series was subject to variation conditioned by laryngeal features. The voiceless stops were lamino-dental (i.e., */t̪/) while their voiced counterparts were more apical alveolar or postalveolar retroflex in articulation (i.e., */ɖ/) (Stampe, 1966, pp. 392, footnote 8; Bhat D. N., 1973, p. 33; Zide, 2008, p. 258).[15] Phonetic variation of this kind is still attested in some Munda languages that lack

---

[15] Widespread variation within Munda led S. Bhattacharya to contemplate a possible three-way contrast between dental, alveolar and retroflex articulations in Proto-Munda, comparable to the system reconstructed for Proto-Dravidian (Bhattacharya, 1975, p. 84). However, there is little to support this point of view and Bhattacharya ultimately retracted it in favour of the one presented here (see Zide, 2008. p. 258).

contrastive retroflexion. For instance, Sora /t/ is described as dental while /d/ is described as alveolar (Ramamurti, 1986 [1933], pp. 66, 266).[16]

Most contemporary Munda languages now distinguish coronal plosives for both voicing and retroflexion (e.g., /t, d/ vs. /ʈ, ɖ/). However, many of these languages exhibit an asymmetry that reflects the phonetic nature of the original system; /t/ and /ɖ/ are frequent in native vocabulary while /d/ and /ʈ/ occur almost exclusively in loanwords. This asymmetry is reported, for instance, in Gorum/Parengi (Aze & Aze, 1973, pp. 217-218; Anderson & Rau, 2008) and in Gutob (Griffiths, 2008).

The eventual phonemic split between dental and retroflex consonants in most Munda languages is generally attributed to areal influences, such as bilingualism in Dravidian and Indo-Aryan languages and the assimilation of numerous loanwords from these sources. Once the phonemic distinction was introduced, the phonetic nature of the original Munda coronal plosives paved the way for their reinterpretation as dental in the case of voiceless allophones and retroflex in the case of voiced allophones.

---

[16] A similar pattern of allophonic variation is reported in some of the non-Munda languages of the Austro-Asiatic family spoken within South Asia. In these cases, however, variation is conditioned by the position of coronal consonants in the word or syllable, not by laryngeal manner. For instance, the Nicobarese languages lack contrastive retroflexion and distinguish only two coronal places of articulation, denti-alveolar and palatal (i.e., /t/ vs. /ʧ/). Nancowry /t/ is described as dental in onsets and alveolar in codas (Radhakrishnan, 1981). In Car Nicobarese, /t/ is described as alveolar word-initially and intervocalically but retroflex word-finally (Das, 1977). These languages show a preference for more anterior articulations in initial (or pre-vocalic) positions and more retracted/posterior articulations in final (or post-vocalic) positions. Thus, it is possible that the phonetic variation of coronal plosives in (pre-)Proto-Munda was conditioned not only by laryngeal features but also by phonotactic position, or some interaction of these two factors.

As peculiar as it may seem, the asymmetry between voicing and coronal place in Munda has been reported elsewhere and may reflect a cross-linguistic phonetic trend that is grounded in articulation. For instance, Dixon notes a correlation between retroflexion and voicing in some Australian languages (2002, p. 571). Hamann & Fuchs (2010) have shown that the articulatory and aerodynamic requirements for voiced dental or alveolar stops can lead to tongue tip retraction and retroflexion. They demonstrate that this pattern occurs synchronically for some speakers of German and suggest that it may be responsible for the diachronic emergence of voiced retroflex plosives (as the sole retroflex plosive) in at least three unrelated languages: Dhao (Malayo-Polynesian), Thulung (Tibeto-Burman), and Afar (East-Cushitic). As discussed in §2.2.1 above, this pattern may be part of a larger trend in which coronal place of articulation is often conditioned to some degree by manner of articulation, with more sonorous manners tending toward more apical posterior articulations.

In sum, evidence suggests that Proto-Munda had the phonemes */t/ and */ɖ/, with a primary "contrastive" opposition for voicing and a secondary "conditioned" opposition for place. That is, retroflexion was conditioned by voicing. The introduction of /d/ and /ʈ/ in Indo-Aryan and Dravidian vocabulary led to a phonemic split between denti-alveolar and retroflex articulations. In the newly restructured system, voiceless Munda */t/ was aligned with the denti-alveolar series while voiced Munda */ɖ/ was aligned with the retroflex series.

2.2.5 Tibeto-Burman

Contrastive retroflexion is an areal innovation in Tibeto-Burman, much as it is in Indo-Aryan and Munda. However, retroflexion is much less extensive in Tibeto-Burman relative to the other families, and has emerged from different historical antecedents. Whereas retroflexion has emerged primarily from post-liquid and post-vocalic coronal consonants in Dravidian and Indo-Aryan, it has emerged in Tibeto-Burman primarily from pre-rhotic consonants, coronal and non-coronal. The examples in (13) demonstrate the correspondence between Cr sequences in Classical Tibetan and retroflex stops in contemporary Central Tibetan. Classical Tibetan reflects the phonological structure of the language in the seventh century CE (DeLancey, 2003a).

(13) Retroflex plosives from Cr- sequences in Tibetan (data from Bhat, 1973, p. 34)

Gloss	Classical Tibetan		Central Tibetan
'child'	phru-gu	>	ʈʰu-gu
'before'	drung-du	>	ʈung-du
'belly'	grod-pa	>	ɖʰö-pa

Developments of the kind illustrated in (13) have taken place in various Tibeto-Burman languages of the western Himalayas (i.e., northwest India and Nepal) and in the Loloish and Qiangic branches, both of which fall outside the confines of South Asia, primarily in China and Myanmar (Matisoff, 2003, pp. 21-23).

The Cr- sequences that produced retroflex consonants in Tibeto-Burman potentially included those with original labial, coronal or velar stops (as shown in (13)), though not all languages derived retroflex consonants from all three types. Bhat (1973) suggests that this development is not the result of /r/ inducing retroflexion in preceding non-coronals. Rather, the initial plosives have been elided, but their laryngeal features have been preserved on the liquid /r/, which has subsequently become /ʈ/ or /ɖ/ (Bhat D. N., 1973, p. 44). Alternatively, it might be possible to regard this as a case of coalescence in which the voicing and manner of the initial C are preserved along with the place of the following liquid.

More often than not, Cr sequences produced retroflex affricates in Tibeto-Burman, or at least plosives with a fricative or rhotic-like release. This appears to be true even in cases where the retroflex consonants in question are represented phonemically as plosives. For instance, regarding the Limi dialect of Humla Bhotia in northwest Nepal, Wilde (2001) writes: “Following the plosives /ʈ, ʈ^{h} and ɖ/ there is a slight fricative-like sound which could be transcribed as [ʈɹ̥, ʈɹ̥^{h} and ɖɹ] respectively” (p. 24). Similarly, Denwood (1999) transcribes the retroflex stops of Central Tibetan as /tr, tr̥, dr/, instead of the more common /ʈ, ʈ^{h}, ɖ/ (as in (13)), and describes them as (apical) alveolar plosives with affrication.

The diachronic development of retroflexion in Tibeto-Burman bears a strong typological resemblance to the development of retroflex affricates in a few Indo-Aryan languages of the Dardic group (see (12) in §2.2.3). In both cases, retroflex consonants derive from original Cr sequences in which C can be coronal or non-coronal, and in both cases the output is primarily an affricated retroflex stop.[17]

In sum, retroflexion has developed in Tibeto-Burman primarily from consonants in pre-rhotic contexts. The consonants occurring in the original Cr sequences that gave rise to

[17] The development of retroflexion from Cr sequences in Tibeto-Burman and Dardic also bears a resemblance to the development of retroflexion from labialized C^{w} consonants in Minto-Nenana (Athapaskan), as described in Hamann (2005). In each of these cases, C can be a non-coronal consonant, the conditiong element (/r/ or labialization) follows C (or at least accompanies its release), and the output is a retroflex segment with affrication.

retroflexion included both coronal and non-coronal stops alike, and the output was typically an affricated retroflex stop.

2.2.6 Summary

The diachronic sources of retroflexion in South Asia are many and varied. Nevertheless, some generalizations can be made. Evidence suggests that the two largest families, Dravidian and Indo-Aryan, have developed retroflex consonants primarily from dental or alveolar consonants in post-liquid (i.e., rt / lt > ʈ) and post-vocalic environments (i.e., ut > uʈ), while Tibeto-Burman languages have developed them primarily from coronal or non-coronal consonants in pre-rhotic contexts (i.e., Cr > ʈ). Munda languages show evidence of retroflexion induced by voicing of denti-alveolar plosives, combined with areal influence from Indo-Aryan and Dravidian. These diachronic developments have not only given rise to retroflex phonemes in South Asia, but also to phonotactic constraints on those phonemes. The phonotactic constraints on retroflex segments in South Asian languages are the subject of the following section.

2.3 Phonotactic restrictions on retroflexion

Cross-linguistically, retroflex consonants are often subject to phonotactic restrictions. That is, their distribution is often limited in comparison to that of their denti-alveolar counterparts. The phonotactic restrictions on retroflexion are of particular importance to the study of retroflex consonant harmony because consonant harmony often has the effect of introducing retroflex consonants in environments where they might not occur otherwise, or at least did not occur historically. This section reviews the phonotactic restrictions on retroflex consonants in South Asian languages. Two broad patterns are identified and discussed. The first pattern, which is characteristic of Dravidian and Old-Indo-Aryan (and possibly Munda), involves a prohibition on retroflex consonants in word-initial position, or positions that are strictly pre-vocalic (§2.3.1). The second pattern, which is characteristic of Tibeto-Burman, involves a prohibition on syllable-final retroflex consonants, i.e., those that are strictly post-vocalic (§2.3.2). These two patterns are contradictory: the positions where retroflex consonants are avoided in one pattern are precisely the positions where they are favoured in the other, and *vice versa*. It is argued here that these contradictory patterns can be explained if phonotactic restrictions on retroflexion are a direct result of the evolution of retroflexion in a given language (§2.3.3).

2.3.1 Restrictions on initial or pre-vocalic retroflex consonants

Contrastive retroflexion is a native feature of Dravidian that can be traced back to Proto-Dravidian. Recall that Proto-Dravidian had a rich coronal inventory that included lamino-dental, apico-alveolar, retroflex and palatal consonants (§2.2.2). However, the distribution of coronal consonants was not symmetrical in Proto-Dravidian. Whereas laminal consonants (dental and palatal) occurred in word-initial, medial and final positions, apical consonants (alveolar and retroflex) did not occur word-initially; they were limited to word-medial and final environments (Zvelebil, 1970, p. 77; Subrahmanyam, 1983, p. 40; Steever, 1998b; Krishnamurti, 2003, pp. 119-120). This distribution is still evident in the few Dravidian languages that preserve the four-way coronal place system of Proto-Dravidian. In the great majority of Dravidian languages, where the alveolar series has merged variously with the dental or retroflex series, the phonotactic restriction on word-initial apicals is still preserved as a restriction on retroflex consonants. Word-initial retroflex consonants are uncommon in the native vocabulary of most Dravidian languages.[18]

The New Indo-Aryan languages tend to exhibit a more symmetrical distribution of coronal consonants, at least when it comes to stops. Dental and retroflex stops typically contrast in word-initial, medial and final environments. Significantly, however, Old Indo-Aryan Sanskrit exhibited an asymmetry much like that in Proto-Dravidian; retroflex consonants "practically never" occurred word-initially in the earliest Sanskrit literature (Masica, 1991, p. 157). The few words with initial retroflex consonants listed in most Sanskrit dictionaries are mostly later innovations that appear in texts only from the fifth and sixth centuries onwards (Jain, 1934, pp. 57–58; Schwarzschild, 1973; Deshpande, 1979). While most New Indo-Aryan languages have extended the dental vs. retroflex contrast to plosives in word-initial position, the prohibition on word-initial retroflexion is still maintained on retroflex sonorants in most (if not all) cases.

The prohibition on word-initial retroflex consonants in Proto-Dravidian and Old Indo-Aryan is not unique to South Asian languages. Australian languages exhibit a very similar pattern. Some Australian languages maintain a four-way coronal place contrast like that of Proto-Dravidian, with two laminal articulations (dental and palatal) and two apical articulations (alveolar and retroflex). Others have a reduced two-way or three-way system with a single laminal or single apical series. The maximum number of coronal place contrasts in any given language is typically found in intervocalic position. In other positions restrictions apply. The

---

[18] Some Dravidian languages have developed word-initial retroflex stops. See §3.1 for discussion.

relevant phonotactic positions can be defined in relation to the template $C_1VC_2C_3VC_4$, in which $\{C_1, C_3\}$ constitute the set of strictly pre-vocalic consonants and $\{C_2, C_4\}$ constitute the set of strictly post-vocalic consonants. Apicals of any kind, alveolar or retroflex, are the least preferred segments in pre-vocalic $\{C_1, C_3\}$ positions and the most preferred segments in post-vocalic $\{C_2, C_4\}$ positions. In other words, they are avoided word-initially and in syllable onsets that are not intervocalic, and favoured in most non-initial positions, which include syllable codas and intervocalic onsets. Some Australian languages, such as Martuthunira, prohibit apicals altogether in $\{C_1, C_3\}$ position. Where apicals do occur in these environments, they are always very infrequent. Languages with a distinction between alveolar and retroflex apicals invariably neutralize the contrast in $\{C_1, C_3\}$ position in favour of one series or the other (Hamilton, 1996; Dixon, 2002).

Traditionally, the phonotactic restriction on retroflexion in Dravidian and Old Indo-Aryan has been defined with respect to word-initial and non-initial positions, not with respect to the more fine-tuned distinction between pre-vocalic $\{C_1, C_3\}$ and post-vocalic $\{C_2, C_4\}$ positions, as in Australian. Nevertheless, the generalization appears to be the same in both cases. The pre-vocalic $\{C_1, C_3\}$ position appears to accurately reflect the environment where retroflexion is prohibited in Proto-Dravidian and Old Indo-Aryan, with the important caveat that retroflex segments can occur in C_3 position if they are part of a homorganic consonant cluster (e.g., -ʈʈ-, -ɳʈ-, etc.), an exception that also applies to Australian.

The status of phonotactic restrictions on retroflexion in Munda (Austro-Asiatic) is much harder to assess. Some languages appear to show a preference for initial dentals where other languages have initial retroflex consonants. In these cases the initial retroflex consonants are arguably later innovations that are the product of retroflex consonant harmony (see §3.4). If so, then Munda might have disfavoured word-initial retroflex consonants at some earlier stage, much like Dravidian and Old Indo-Aryan. However, it is not clear to what extent this pattern reflects the native Munda system, or to what extent it reflects the influence of Dravidian and Indo-Aryan on that system (but cf. footnote 16 on p. 81).

In sum, Dravidian and Old Indo-Aryan both avoided retroflexion in word-initial and other strictly pre-vocalic positions. The same trend is also attested in Australian languages. However, many Tibeto-Burman languages of South Asia exhibit a very different phonotactic restriction. This is the subject of the following section (§2.3.2).

2.3.2 Restrictions on final or post-vocalic retroflex consonants

Like Dravidian and Old Indo-Aryan, some Tibeto-Burman languages of South Asia also exhibit phonotactic restrictions on retroflexion. However, the Tibeto-Burman pattern is precisely the inverse of the Dravidian and Old Indo-Aryan pattern. Whereas Dravidian, Old Indo-Aryan and Australian avoid initial or strictly pre-vocalic retroflex segments, Tibeto-Burman languages prohibit final or strictly post-vocalic retroflex segments.

A prime example of the Tibeto-Burman pattern is found in Lhomi, a language of northeastern Nepal. Lhomi distinguishes the retroflex plosives /ʈ, ʈʰ/ from their dental counterparts /t, tʰ/. Both dental and retroflex consonants occur in syllable onsets, regardless of whether the onsets are word-initial, intervocalic or post-consonantal. However, the retroflex series does not occur in syllable codas. Only coronals of the dental series occur in that environment, together with labials and velars. In terms of the $C_1VC_2C_3VC_4$ template introduced earlier, the retroflex consonants of Lhomi occur in pre-vocalic $\{C_1, C_3\}$ position (in addition to intervocalic position), but not in the strictly post-vocalic {C_2, C_4} position. Representative examples are listed in (14) and (15) (Vesalainen & Vesalainen, 1976).

(14) Dental and retroflex plosives in Lhomi syllable onsets (i.e., pre-vocalic position)

Dental /t/		Retroflex /ʈ/	
tá	‘horse’	ʈák	‘button’
sóp.tok	‘ring’	sìp.ʈok	‘comb of chicken’
sà.tu	‘to eat’	pʰí.ʈa	‘wild cat’

(15) No retroflex consonants in Lhomi syllable codas (i.e., strictly post-vocalic position)

Dental /t/		Retroflex /ʈ/
pèt	‘is’	*pèʈ
sórit	‘stomach (hon.)’	*sóriʈ
lít.maŋ	‘he didn’t come’	*líʈ.maŋ

The phonotactic restriction on retroflexion in Lhomi is attested in many other Tibeto-Burman languages of the western Himalayas including, Tshangla (Andvik, 2003), Nar Phu (Noonan, 2003b), Tamang (Mazaudon, 2003), Dolpo (Kevin Kopp, p.c.), Humla (Wilde, 2001), and Dolakha Newar (Genetti, 2007), among others.[19]

---

[19] In Lhomi, dental /t/ has a glottal [ʔ] allophone in syllable codas. This is not uncommon in Tibeto-Burman, where coda consonants are often subject to glottalization or, in some cases, complete elision. This obscures the distribution pattern to some degree because both dental and retroflex consonants may be absent in syllable codas at a phonetic level in some Tibeto-Burman languages. Nevertheless, language-internal and historical-comparative evidence in these languages support the dental series in syllable codas, either phonemically or historically, but not the retroflex series.

In sum, Tibeto-Burman languages tend to prohibit retroflex segments in syllable-codas where they are strictly post-vocalic. This phonotactic pattern is precisely the inverse of that exhibited by Dravidian and Old Indo-Aryan, where retroflex segments are preferred in post-vocalic positions and avoided in word-initial or pre-vocalic positions.

2.3.3 Diachronic and perceptual bases of retroflex phonotactics

The phonotactic restrictions on retroflexion in a given language family derive more or less directly from the particular historical development of retroflexion in that family, though perceptual factors may also play a role. Dravidian, Indo-Aryan and Australian are all language families in which retroflexion has developed from dental or alveolar coronals primarily in post-liquid and/or post-vocalic positions. Post-liquid and post-vocalic consonants are by definition non-word-initial, at least when they are tautomorphemic with the preceding liquid or vowel. Retroflexion in a post-liquid environment typically entails loss of the conditioning liquid (e.g., rt, lt → ʈ). This could conceivably introduce word-initial retroflex segments in a language if sequences such as /rt/ or /lt/ occurred word-initially. As it is, such consonant clusters do not occur initially in most languages because they violate the sonority sequencing required for well-formed onsets. Thus, the asymmetrical distribution of retroflex consonants in Dravidian, Indo-Aryan and Australian can be viewed as a natural consequence of their historical development; they are limited to non-initial positions because their historical antecedents were all non-initial (i.e., -lt, -rt, -ut, etc.).

The Tibeto-Burman languages provide strong support for this conclusion. Recall that the retroflex consonants of Tibeto-Burman have developed primarily from non-retroflex consonants in pre-liquid positions. Unlike rC clusters with post-liquid consonants, Cr clusters with pre-liquid consonants can and do occur frequently in word-initial position and in other syllable onsets, but are not common in word-final position or in syllable codas of any kind. Thus, it is not surprising to find that the retroflex consonants of Tibeto-Burman, which derive from original Cr sequences, are restricted to syllable onsets and prohibited in syllable codas. Once again, the limited distribution of retroflex segments in these languages is a natural consequence of their historical development; they are limited to syllable onsets because their historical antecedents occurred only in that position.

The prohibition on retroflex consonants in word-initial or other strictly pre-vocalic contexts, as exemplified in Dravidian, Old Indo-Aryan and Australian languages, is probably the most common phonotactic restriction on retroflexion cross-linguistically. Several studies have suggested that this restriction is grounded in speech perception (Hamilton, 1996; Steriade, 2001; Hamann, 2003). Recall that the acoustic cues to retroflexion are most perceptually salient in VC transitions and least salient in CV transitions (§1.2.3). Thus, it is argued that contrastive retroflexion is restricted to post-vocalic VC contexts where the cues that signal retroflexion are most robust. Contrastive retroflexion is avoided or neutralized in strictly pre-vocalic positions because the CV transitions of retroflex and denti-alveolar coronals are perceptually similar; contrast between the two is difficult to maintain without the benefit of the more salient cues that reside in VC transitions.

The perceptual account is not meant to predict that retroflex consonants cannot be distinguished from denti-alveolar coronals in word-initial or other strictly pre-vocalic contexts. However, it has been argued to predict an implicational universal that applies to apical subtypes; a language is expected to maintain contrast between apico-alveolar and retroflex segments in perceptually weak CV environments only if it also maintains it in perceptually salient VC contexts. This is reflected in Steriade's (2001) law of apical contrast in (16).

(16) Patterns of apical contrast and neutralization (Steriade, 2001, p. 226)[20]

a. **The Law:** if the t/ʈ contrast occurs in a language, it occurs after V

b. **The General Case:** t/ʈ contrast only after V

c. **The Initial Deviation:** t/ʈ contrast only after V and in #__

Steriade's law of apical contrast in (16) predicts that no language should exhibit contrast between retroflex and non-retroflex apicals in word-initial (or other strictly pre-vocalic) CV contexts without also maintaining the contrast in strictly post-vocalic VC contexts. Steriade's law appears to hold true, but only as it applies to the contrast between apical subtypes; that is, the contrast between apico-alveolar and retroflex segments in languages that distinguish both of these from laminal segments, as in some Australian and South Dravidian languages (i.e., /t̻, t̺, ʈ/).[21]

---

[20] Steriade (2001) also notes another deviation from the general pattern in (16)(b), which she calls "The I-Deviation". In this deviation, /t/ and /ʈ/ contrast after central and back vowels but are neutralized to [t] after [i]. This pattern is the result of a general articulatory incompatibility between retroflexion on the one hand and front vowels or palatalization on the other (cf. Hamann 2003).

[21] It is not clear whether Steriade (2001) intended the laws concerning apical contrast and neutralization (in (16)) and inter-apical assimilation (to be discussed in §2.4, see (17)) to apply only to apical segments in the narrowest sense (i.e., the class of apico-alveolar and apico-retroflex segments, excluding laminal denti-alveolars), or whether they are intended to apply to retroflex and denti-alveolar segments more broadly (i.e., including laminal denti-alveolars). On the one hand, most of the wording in her paper suggests the narrower interpretation. On the other hand, she cites examples from Indo-Aryan and Dravidian languages in which the relevant class of segments is clearly retroflex and lamino-dental, not retroflex and apico-alveolar (at least in the case of plosives; e.g., Sanskrit, Panjabi, and other languages cited in the Appendix to her paper). The inclusion of these examples suggests a broader interpretation. Whatever the case may be, the laws (and their qualifications) appear to hold with few or no

Significantly, Steriade's law cannot be extended to account for patterns of contrast and neutralization between retroflex and non-retroflex anterior coronals in general. This is abundantly clear in light of the evidence from Tibeto-Burman languages such as Lhomi in (14) and (15) above. If Steriade's law of apical contrast were interpreted as a general law governing contrastive retroflexion, then it would predict that languages such as Lhomi should not exist. This is because Lhomi fails to maintain contrastive retroflexion in precisely those environments where the law in (16) would predict that it should occur (i.e., VC contexts); and it maintains contrastive retroflexion exclusively in those environments where the law in (16) would predict that it should not occur, or where it should occur only as a last resort.

In conclusion, the evidence from South Asian languages suggests that the phonotactic restrictions on retroflexion in a given language arise more or less directly from the particular diachronic developments that produced retroflexion in that language. No doubt perceptual factors also play a role. However, the role of perceptual factors is only indirect, in so far as they influence the historical development of retroflexion. For instance, the salience of retroflex place cues in VC transitions may account for the overall prevalence of retroflexion in post-liquid environments cross-linguistically. These are precisely the environments where any retroflex-like cues produced by the liquid could easily be misparsed and attributed to the following coronal consonant. Conversely, since retroflex cues are less salient in CV transitions, retroflexion is expected to occur less frequently in Cr clusters. Thus, while perceptual factors

---

exceptions when interpreted in the narrow sense, but hold only as generalizations when interpreted in the broader sense.

may account for the cross-linguistic prevalence of certain diachronic developments over others, it is ultimately the diachronic developments themselves that give rise to phonotactic constraints on retroflexion, not the perceptual factors.

The phonotactic restrictions on retroflexion reviewed in this section are of particular importance to the study of retroflex consonant harmony. This is because consonant harmony often has the effect of introducing retroflex consonants in environments where they did not occur historically, most notably in word-initial position in the case of Dravidian, Indo-Aryan and Munda. Before turning to the survey of retroflex consonant harmony in South Asian languages, it is useful to review some of the unique properties exhibited by retroflex segments in patterns of local coronal assimilation.

2.4 Retroflexion in local assimilation

Cross-linguistically, major place assimilation between adjacent consonants is predominantly regressive. In a -C_1C_2- sequence, C_1 is typically the target of assimilation while C_2 is the trigger. As a result, the output of local assimilation in a -C_1C_2- sequence is typically -C_2C_2-, not -C_1C_1- (Jun, 2004). However, patterns of minor coronal place assimilation constitute an important exception to this trend, particularly when retroflexion is involved. Unlike major place features/gestures, retroflexion has a strong tendency to trigger progressive assimilation. Thus, in a -C_1C_2- sequence, where C_1 is retroflex and C_2 is a denti-alveolar coronal, the output of local assimilation is typically -C_1C_1-, not -C_2C_2-. This trend is reflected in Steriade's (2001) law of inter-apical assimilation, which is summarized in (17).

(17) Patterns of inter-apical assimilation (Steriade, 2001, p. 227)

a. **The Law:** all else equal, assimilation is progressive in apical clusters

b. **Final Deviation:** assimilation may be regressive across the boundary of content words

c. **Nasal Deviation:** assimilation may be regressive in nasal-stop clusters

According to the law of inter-apical assimilation in (17)(a), assimilation between two adjacent apicals is predominantly progressive (i.e., [ʈt] → [ʈʈ], [tʈ] → [tt]). Steriade argues that assimilation can be regarded as “perceptually tolerated articulatory simplification” (2001, p. 232). Assimilation for a feature is motivated by articulatory simplification, but it targets those positions in which contrast for the feature is least salient and, therefore, least likely to be missed if it is neutralized. Recall that major place cues are most salient in CV transitions, whereas apical/retroflex cues are most salient in VC transitions (§1.2.3). This means that, in the case of major place assimilation, C_1 is the most likely target of assimilation because it lacks critical CV cues, whereas C_2 is the most likely trigger because it preserves those cues. As a result, major place assimilation is predominantly regressive. In the case of apical assimilation, however, the distribution of cues is reversed; C_2 lacks the crucial VC cues while C_1 maintains them. As a result, the directional trend is reversed and assimilation is predominantly progressive for apicals.

In (17)(b)–(c), Steriade notes two important exceptions to the trend favouring progressive assimilation among apicals. First, apical assimilation may be regressive across word boundaries. This can be attributed to positional faithfulness. That is, word-initial consonants are more likely to be preserved whether they are poorly cued or not. Secondly, apical assimilation may be regressive in nasal + stop sequences. This can be attributed to independent factors concerning nasals. Nasals are known to have less salient place cues than other manners of articulation. Therefore they are the most preferred targets of consonantal place assimilation cross-linguistically (Jun, 2004).

Steriade's law of inter-apical assimilation is most reliable when interpreted in the narrow sense as a law governing assimilation between apical subtypes, i.e., the class of apico-alveolar and retroflex segments, excluding laminal denti-alveolars. Applied to the larger class of retroflex and denti-alveolar segments, broadly defined to include apical and laminal articulations, the law may hold only as a generalization.

Examples of regressive retroflex assimilation targeting dental segments are not difficult to find. In fact, Bhat (1973) reached the conclusion that local retroflex assimilation was primarily anticipatory and regressive. Many of the cases cited by Bhat can be explained in terms of the systematic exceptions to progressive apical assimilation in (17)(b)–(c) (i.e., nasal place assimilation and the preservation of word-initial segments). However, it remains to be seen whether all cases of regressive retroflex assimilation can be explained in terms of these or other systematic exceptions. In the interim, it is best to follow Hamann (2003) in assuming a softer interpretation of Steriade's law. According to Hamann, "there are perceptual reasons why retroflexion should spread preferably towards the following segment, but these motives are not as strong in every language as to result in a universal pattern of progressive assimilation for retroflexes" (2003, p. 124).

Apart from the perceptual factors, there are other factors that might contribute to the trend toward progressive retroflex assimilation. For instance, the bias toward progressive assimilation might stem partly from diachronic developments. Recall that retroflex segments develop most often from denti-alveolar coronals in post-liquid environments (§2.2), and that this development ultimately produces an asymmetrical distribution of coronal consonants, such that retroflex segments are absent in word-initial position (§2.3). In systems of this kind (which may constitute a majority of the world's languages with retroflexion), retroflex segments are relatively common in morpheme-final position, but not in morpheme-initial position. This is especially true in the case of inflectional and derivational affixes, which typically constitute a small, closed and phonologically conservative set of morphemes. It follows from this that sequences such as ʈ + t, with a morpheme-final retroflex followed by a morpheme-initial denti-alveolar, should arise more or less frequently in the course of inflection and derivation. Such sequences set the stage for progressive retroflex assimilation (i.e., ʈ + t → ʈʈ). However, sequences such as t + ʈ, with a morpheme-final denti-alveolar followed by a morpheme-initial retroflex, are not expected at all. Thus, the prevalence of progressive assimilation over regressive assimilation might simply reflect the fact that the conditions for regressive assimilation rarely arise.

Finally, it is worth noting that retroflex segments tend to dominate other coronals in patterns of local assimilation. That is, in local coronal assimilation, retroflex segments are almost always the triggers of assimilation but rarely the targets, whereas denti-alveolars are almost always the targets but rarely the triggers. Steriade (2001, p. 227) notes that there are cases in which a denti-alveolar segment triggers progressive assimilation (i.e., de-retroflexion) in a following retroflex segment (i.e., /tʈ/ → [tt]). On the basis of this observation, she argues that the trend toward progressive apical assimilation cannot be reduced to a retroflex dominance effect in which retroflex segments always trigger assimilation in adjacent denti-alveolars regardless of their position in a consonant cluster (i.e., /ʈt/, /tʈ/ → [ʈʈ]). However, cases of progressive de-retroflexion of the kind cited by Steriade are exceedingly rare and, as we have already noted, retroflex assimilation can be progressive and/or regressive. Thus, it is possible that some languages do exhibit retroflex dominance effects, even if some do not.

Palatals also tend to dominate denti-alveolars in patterns of local coronal assimilation but evidence bearing on dominance relations between palatal and retroflex segments is scarce. In most cases the relevant sequences simply fail to arise for phonotactic reasons, or, if they do, they fail to show assimilation of any kind. Where assimilation does occur, no clear cross-linguistic trend is evident. For instance, Polish retroflex sibilants assimilate to a following palatal under conditions of local assimilation (e.g., /ʐ+ʑ/ → [ʑʑ]; /ʂtɕ/ → [ɕtɕ] (Jarmasz, 2008, p. 27, citing data from Dyszak, 1997)), but palatal sibilants assimilate to a following retroflex under conditions of non-local assimilation in Indo-Aryan languages such as Kalasha and Indus Kohistani (e.g., /ʃ…ʂ/ → /ʂ…ʂ/) (§3.3).

In sum, retroflex assimilation is unique in relation to other kinds of local assimilation. Whereas major place assimilation between adjacent consonants is predominantly regressive, local retroflex assimilation is predominantly progressive, or at least bi-directional. This directional asymmetry can be attributed to the asymmetrical distribution of perceptual cues to place contrast. However, other factors may also play a role including: (i) historically-motivated phonotactic restrictions that favour the conditions necessary for progressive assimilation over those necessary for regressive assimilation; and (ii) a general dominance of retroflex and palatal articulations over denti-alveolar articulations in patterns of coronal assimilation.

2.5 Summary and conclusions

In this chapter we have seen that contrastive retroflexion is a widespread areal feature of South Asia, one that occurs in most languages of the region regardless of their genetic affiliation. The predominant coronal place system in South Asia is one with three distinct places of articulation: dental (or denti-alveolar), retroflex and palatal. Contrastive retroflexion occurs in all manners of articulation but is most frequent among plosives, less frequent among nasals, liquids and fricatives, and least frequent among approximants and vowels.

In Dravidian, Indo-Aryan and Tibeto-Burman, retroflexion originated phonetically in the class of liquids and rhotic approximants and spread via local or non-local assimilation from these segments to other manners of articulation, where it ultimately took on phonemic status.[22] In the case of Dravidian and Indo-Aryan, retroflexion spread progressively targeting anterior coronals in post-liquid environments (e.g., -lC-, -rC-, etc., but also in post-vocalic environments such as -uC- in OIA); in the case of Tibeto-Burman, retroflexion spread regressively, targeting consonants in pre-liquid environments (i.e., Cr-). These diachronic developments yielded phonotactic restrictions (i.e., asymmetrical distributions) for retroflex consonants in each

[22] It is worth noting yet another asymmetry here: the class of segments that are the first to develop phonetic retroflexion (e.g., liquids and rhotic approximants) are among the least likely to maintain contrastive retroflexion, whereas the class of segments that are the last to develop retroflexion (e.g., plosives) are the most likely to maintain it as a contrastive feature. This trend may seem counter-intuitive. For instance, we might expect contrastive retroflexion to occur more frequently where it is also more natural phonetically. However, the asymmetry can be explained from the perspective of contrast. It is precisely because liquids and approximants are naturally inclined toward phonetic retroflexion that they make poor candidates for contrastive retroflexion. The situation is analagous to that of voicing in sonorants. Contrastive voicing is rare among sonorants because sonorants are naturally inclined toward phonetic voicing. The naturalness of apicality and retroflexion in the class of liquids and rhotic approximants makes their occurrence there somewhat redundant. Thus, retroflexion is more likely to be contrastive in the class of plosives where its occurrence is less predictable.

family. Retroflexion in post-liquid and post-vocalic environments yielded systems with no initial retroflex segments in Proto-Dravidian and Old Indo-Aryan, whereas retroflexion in pre-liquid environments yielded a system with no syllable-final retroflex segments in Tibeto-Burman. The distribution of perceptual cues to retroflexion may also play a role in conditioning some of these phonotactic restrictions, at least to the extent that they have influenced the diachronic developments or reinforced the patterns that emerged from them.

Retroflexion exhibits a unique directional asymmetry in patterns of local assimilation. Whereas major place assimilation between adjacent consonants is predominantly regressive, local retroflex assimilation is predominantly progressive, or at least bi-directional. As in the case of phonotactic restrictions on retroflexion, the asymmetry can be attributed to both historical and perceptual factors, both of which are probably inter-related. The tendency toward bi-directional assimilation may also be partly attributed to a dominance effect in which retroflex segments dominate other coronals in assimilation, particularly anterior denti-alveolars.

Local retroflex assimilation is very common cross-linguistically. It applies more often than not wherever retroflex segments come into contact with their non-retroflex denti-alveolar counterparts. However, in her cross-linguistic study of retroflexion, Hamann (2003) found very few cases of non-local retroflex assimilation and was forced to conclude that "Long-distance assimilation is observed *very infrequently* in languages with retroflexes" (p. 124, emphasis mine). The following chapter presents a survey of long-distance retroflex assimilation in South Asia and demonstrates that, contrary to previous conclusions, it is very common and widespread, although it has often gone unnoticed, or at least unreported, in the literature.

Chapter 3
A survey of retroflex consonant harmony in South Asia

Retroflex consonant harmony has received little attention in the literature on South Asian languages. References to it are few and brief. No detailed studies have been offered, either with respect to individual languages or with respect to language families, and it has often gone unnoticed, or at least unreported, in many languages. From these things it would be possible to infer that retroflex consonant harmony is a rare and insignificant phenomenon in South Asia. This is not the case. The current study, which is the first of its kind for the region, presents evidence that retroflex consonant harmony is a widespread areal trait affecting a large number of languages in the northern half of the South Asian subcontinent, including languages from at least three of the four major South Asian families: Dravidian, Indo-Aryan and Munda.

The study draws on data from a large body of published and unpublished sources including dictionaries, vocabulary lists, grammars and electronic databases. It demonstrates that retroflex consonant harmony is widespread in South Asia as a synchronic morpheme structure constraint (MSC) that is the product of diachronic assimilation. Historically, many South Asian languages, particularly those of the Dravidian and Indo-Aryan families, did not have word-initial retroflex consonants (for reasons discussed in §2.3). As a result, roots containing two non-adjacent coronal consonants were limited to just two of four possible configurations involving dental and retroflex segments: dental-dental (T-T) and dental-retroflex (T-Ṭ), but not retroflex-dental (Ṭ-T) or retroflex-retroflex (Ṭ-Ṭ). This is represented schematically in (1)(a).[1]

(1) Two co-occurrence patterns affecting dental (T) and retroflex (Ṭ) plosives

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

Many South Asian languages, which formerly had co-occurrence patterns like that in (1)(a), have developed consonant harmony systems like that in (1)(b), in which co-occurring coronal consonants must agree with respect to retroflexion or non-retroflexion. The pattern in (1)(b) has developed from (1)(a) through a process of regressive retroflex assimilation, in which roots of the type T-Ṭ have become Ṭ-Ṭ. The study employs two kinds of evidence to reveal this pattern: (i) statistical evidence of synchronic co-occurrence restrictions on retroflex consonants in the roots/words of various languages and (ii) historical and comparative evidence demonstrating that roots/words with retroflex consonant harmony (Ṭ-Ṭ) can often be traced to disharmonic cognates (T-Ṭ) in a parent language and/or in closely related languages or dialects.

Each of the South Asian language families is discussed in turn beginning with Dravidian (§3.1), followed by Indo-Aryan (§3.2), Dardic and Burushaski (§3.3), Munda (§3.4)

---

[1] Here, and elsewhere, T stands for any dental plosive and Ṭ for any apical/retroflex plosive. The notation “T-T” stands for a sequence of two consonants of the specified type that are non-adjacent (i.e., separated minimally by a vowel). An asterisk (*) indicates that a sequence is prohibited; a check mark (✓) indicates that it is unrestricted. Other shorthand conventions used throughout the dissertation include: S for any dental/alveolar fricative; Š for any laminal postalveolar ‘palatal’ fricative; Ṣ for any retroflex fricative; TS for any dental/alveolar affricate; Č for any laminal postalveolar ‘palatal’ affricate; Ç for any retroflex affricate; and Ṛ for any retroflex sonorant.

and Tibeto-Burman (§3.5). Although they are classified as Indo-Aryan, the Dardic languages of northern Pakistan are presented separately in §3.3 because they exhibit some unique properties that warrant independent discussion. The isolate Burushaski is included with them because of its geographic proximity and typological relation to the Dardic group. For each language family or subgroup (with the exception of Munda) one or two detailed case studies are presented to establish the presence (or absence) of retroflex consonant harmony and to highlight any typological properties that they might exhibit. Where appropriate an attempt is then made to establish the full extent of retroflex consonant harmony within a family or subgroup by examining, in less detail, the co-occurrence restrictions on retroflex consonants in a broad sample of languages from that group. Finally, a summary of the extent of retroflex consonant harmony in South Asia is provided, along with a discussion of its typological properties (§3.6).

3.1 Dravidian

Proto-Dravidian distinguished dental consonants from apicals, both alveolar and retroflex (§2.2.2). The apicals were subject to a phonotactic restriction: they did not occur word-initially (§2.3.1). As a result, the co-occurrence of dental and apical plosives in Proto-Dravidian roots was limited to just two of four possible configurations. This is represented schematically in (2).

(2) Co-occurrence of dental (T) and apical (Ṭ) plosives in Proto-Dravidian roots [2]

Initial dental:	✓T-T	✓T-Ṭ
Initial apical:	*Ṭ-T	*Ṭ-Ṭ

---

[2] In the case of Proto-Dravidian, the class of apical plosives represented by Ṭ includes both apical alveolar and retroflex plosives. Elsewhere, the class of apical plosives represented by Ṭ is limited to retroflexes.

As shown in (2), all Proto-Dravidian roots containing two coronal plosives had initial dentals, not initial apicals (whether alveolar or retroflex). Only a few South Dravidian languages preserve the apical-alveolar series of Proto-Dravidian, but all contemporary Dravidian languages, without exception, preserve the contrast between dental and retroflex stops. Moreover, most Dravidian languages have inherited the restriction on initial apicals as a restriction on initial retroflexes. In the majority of Dravidian languages, word-initial retroflex consonants are rare or absent altogether in native vocabulary.

Despite the original prohibition on word-initial apicals, some Dravidian languages have developed word-initial retroflex stops from a variety of sources. The most commonly cited sources are: (i) onomatopoeic words, which tend to favour retroflex segments in South Asian languages (e.g., Tamil /ʈaɳɳen̪al/ 'the sound of a bell'); (ii) loanwords from Indo-Aryan and English (e.g., Tamil /ʈaːvuɳ/ < Eng. 'town');[3] and (iii) a rule of metathesis known as 'apical displacement', which affected the South-Central Dravidian group (e.g., compare Gondi /ɖig-/ 'to descend' with Gadaba /iɖg-/) (Zvelebil, 1970, p. 102; Subrahmanyam, 1983, pp. 225–248; Steever, 1998b, pp. 16–17; Krishnamurti, 2003, pp. 157–163).

Another source of word-initial retroflex plosives in Dravidian – one that has received much less attention – is a process of regressive retroflex consonant harmony; initial dental plosives have become retroflex under the influence of a following non-adjacent retroflex plosive within the same root (Burrow & Bhattacharya, 1963, p. 240; Subrahmanyam, 1983, pp. 361–362). Schematically, roots of the type T-Ṭ have become Ṭ-Ṭ in some languages. As a

---

[3] The alveolar plosives of English (including those that occur word-initially) are adapted as retroflex in most South Asian languages (Jagannath, 1981; Koshal, 1978; Ohala, 1978).

result, many of the Dravidian languages that admit retroflex plosives word-initially in native non-onomatopoeic vocabulary also tend to exhibit the coronal co-occurrence pattern in (3).

(3) Co-occurrence of dental (T) and retroflex (Ṭ) plosives in languages with retroflex consonant harmony (*T-Ṭ → Ṭ-Ṭ)

Initial dental:	✓T-T	*T-Ṭ
Initial retroflex:	*Ṭ-T	✓Ṭ-Ṭ

The following subsections present case studies of Malto (§3.1.1), a North Dravidian language, and Pengo (§3.1.2), a South-Central Dravidian language. The studies demonstrate that retroflex consonant harmony has applied in these languages, resulting in a co-occurrence pattern like that in (3). Section §3.1.3 provides evidence that this pattern is not unique to Malto and Pengo but is characteristic of most North and South-Central Dravidian languages, and also of Central Dravidian Parji. Historical-comparative data supporting this conclusion are presented in §3.1.4. Finally, sections §3.1.5 and §3.1.6 examine the roles played by manner of articulation and laryngeal features in retroflex consonant harmony. There, it is argued that manner is a significant conditioning factor, but laryngeal features are not; harmony holds between coronal plosives regardless of laryngeal distinctions, but not between plosives and sonorants even when retroflexion is contrastive in both manner classes.

3.1.1 Malto (North Dravidian)

One of the best known cases of consonant harmony in South Asia is that of Malto, a North Dravidian language spoken in the Rajmahal hills of Jharkhand state in north India. Consonant harmony was first reported in Malto by Mahapatra (1979). It has also received attention in subsequent descriptions of the language (Steever, 1998d) and in recent cross-linguistic surveys of consonant harmony systems (Hansson, 2001; 2010; Rose & Walker, 2004). Mahapatra (1979) reports at least three patterns of consonant harmony in Malto: retroflex harmony, dorsal harmony, and a pattern of lateral harmony, which is restricted to the Sawriya dialect (i.e., *n…l → l…l).[4] The present study is concerned only with the pattern of retroflex harmony. For details concerning the others see Mahapatra (1979, pp. 38–39, 207), Hansson (2001, p. 97) and Appendix B. The consonant phonemes of Malto are listed in (4).

(4) Consonant phonemes of Malto (Mahapatra, 1979; Steever, 1998d)

LAB	DEN	ALV	RET	PAL	VEL	UVL	GLOT
p	t		ʈ	ʧ	k	q	
b	d		ɖ	ʤ	ɡ		
	ð	s				ʁ	h
m	n			ɲ	ŋ		
		l					
		r	ɽ				
w				j			

As shown in (4), Malto distinguishes dental and retroflex plosives. It also has a single retroflex sonorant, the flap /ɽ/, which is distinguished from alveolar /r/. Malto is among those Dravidian languages that have developed word-initial retroflex plosives. Dental and retroflex plosives contrast initially and non-initially in roots containing a single coronal plosive (e.g.,

[4] Mahapatra (1979, pp. 39–40) also describes a laryngeal co-occurrence restriction in Malto. A CVC syllable with an initial voiced velar stop cannot have a voiceless velar coda and one with an initial voiced palatal stop cannot have a voiceless palatal coda. Thus, *ɡVk and *ʤVʧ do not occur but all other voicing combinations are unrestricted (i.e., ɡVɡ, kVɡ, kVk, ʤVʤ, ʧVʤ, ʧVʧ). This can be construed as a kind of directional voicing harmony on homorganic combinations of velar and palatal stops (cf. Hansson 2001, p. 154). However, it may also stem partly from that fact that the native Dravidian phonological system lacked initial voiced stops historically.

/topa/ ‘bunch’ vs. /ʈopa/ ‘drop of liquid’; /pati/ ‘sharp’ vs. /paʈi/ ‘domestic animals which have not given birth’). Roots containing two coronal plosives are more restricted. According to Mahapatra (1979), dental and retroflex plosives do not co-occur within a root. If there are two coronal plosives in a root, then both are either dental or retroflex, as shown in (5).

(5) Retroflex consonant harmony in Malto (data from Mahapatra, 1979, 1987)

a. Harmony across intervening -V-

tot-	‘to hurry’	ʈeʈu	‘hand’
ta:to	‘weaver’	ʈu:ɖ	‘tiger’
doti	‘men’s wear’	ɖeʈa	‘corn cob’
dudo	‘name of tree’	ɖaɖe	‘forest’

b. Harmony across intervening -VC-

tind-	‘to feed’	ʈunɖ-[5]	‘to see’
dundo	‘owl’	ɖanɖa	‘staff’
taɽte	‘tongue’	ɖebɖe	‘crooked’

c. No harmony across an intervening morpheme boundary

taŋ-ɖo	‘his brother’
kaʈ-tan	‘I crossed’
ʈu:ɖ-du	‘tiger’ (with formative suffix /-du/)

[5] In South Asian languages, coronal nasals tend to be homorganic with following coronal stops. By convention, they are often represented as /n/ whenever their precise phonetic realization has no independent phonemic status. Thus, /nɖ/ typically corresponds to phonetic [ɳɖ] in the literature on South Asian languages. This is true of Mahapatra’s (1979, 1987) data and of other language data cited throughout this chapter.

The co-occurrence pattern in (5) fits the description of retroflex consonant harmony. Non-adjacent coronal consonants in a root must agree for retroflexion or non-retroflexion. Retroflex harmony holds between two coronal plosives in a root regardless of whether they are separated by an intervening vowel, as in (5)(a), or an intervening vowel and consonant, as in (5)(b). Harmony does not hold across morpheme boundaries, as exemplified in (5)(c).

Mahapatra's (1979) observation can be verified by applying the statistical methods introduced in §1.4 to lexical data from his (1987) Malto dictionary. To this end, a count was made of all headwords in the dictionary containing word-initial $\#C_1V(N)C_2$ sequences in which C_1 and C_2 are both coronal plosives or retroflex sonorants. Words containing an intervening homorganic nasal (N) were also included in the count. Before presenting the results of the study, it will be useful to justify this particular selection of data because it applies not only to the study of Malto, but also to most other case studies presented in this chapter.

First of all, the count is based on dictionary entries or 'headwords'. Ideally, it might be preferable to count unique roots in a language. By counting headwords there is the likelihood that some C_1-C_2 configurations will receive more counts than they would otherwise. This is because a given root may appear in more than one headword if it is used in forming compounds, or if the dictionary lists derived and inflected forms of the root as independent entries. However, a count based on unique roots presupposes a thorough morphological analysis of the vocabulary. In the absence of such an analysis, a simple count of headwords can suffice as a coarse-grained approximation of the surface pattern. This is because, leaving compounds aside, it is safe to assume that if the dictionary lists derived and inflected forms of roots as independent headwords, it does so for all types of roots. That is, counts of roots with and without consonant harmony are both multiplied by the inclusion of derived and inflected forms leaving the relative frequency of each type more or less intact.

Secondly, the count is limited to word-initial $\#C_1V(N)C_2$ sequences. Ideally, it might be preferable to examine C_1-C_2 sequences over any distance. However, the evidence from Malto in (5)(c) combined with a general lack of evidence for alternations in languages of the region suggests that, where it does occur, retroflex harmony occurs primarily as a static morpheme structure constraint. Thus, it is desirable to limit the count to sequences in which both C_1 and C_2 occur within the same morpheme. Once again, this presupposes a thorough morphological analysis of the vocabulary in order to determine which sequences are morpheme-internal and which ones are not. In the absence of such an analysis the number of potential morphologically complex items can be reduced by restricting the count to word-initial $\#C_1V(N)C_2$ sequences. The greater the distance between C_1 and C_2, the greater the chances that they belong to different morphemes, since most morphemes generally consist of no more than one or two syllables. In addition, the closer the sequence is to the word-initial position, the greater the chances are that it belongs to the root morpheme since South Asian languages are predominantly head-initial in morphological structure; suffixes are the norm and prefixes are relatively rare. By limiting the count to word-initial $\#C_1V(N)C_2$ sequences the study approximates a count of root morphemes without the need for an in-depth morphological analysis of every language.

Finally, a few words are in order about the inclusion or exclusion of nasals in the counts. In South Asian languages, nasals are predominantly homorganic with following stops (plosive or affricate), and homorganic NC clusters occur frequently in roots. The predictable nature of nasal place in NC clusters raises some issues for the count of long-distance co-occurrence patterns when it comes to words containing sequences of $\#C_1VNC_2$ in which N is homorganic with C_2. One option would be to limit the count to consonants separated only by a vowel. In this case a word such as Malto /tind-/ 'to feed' would receive a single count as an instance of /t…n/. Another option would be to count each member of the NC cluster independently. In this case /tind-/ would be counted twice: once as an instance of the sequence /t…n/ and once as an instance of /t…d/. Both of these options can produce misleading results. They can suggest a stronger relation between C_1 and N than warranted, given that the place of N is conditioned primarily by C_2, not necessarily by any restriction in relation to C_1. For this reason I have adopted a third option: $\#C_1VNC_2$ sequences are included in the counts, but in these cases only the combination of C_1 and C_2 is counted. Thus, an example such as /tind-/ is counted only as an instance of /t…d/, not as an instance of /t…n/. This means that all counts of C_1-C_2 sequences, in which C_2 is a coronal nasal, are restricted to cases where the nasal occurs *without* a following coronal stop to condition its place of articulation (e.g., Malto /tan/ 'if'; /d̪aːni/ 'wife, mistress'; etc.).

With these things in mind we return to Malto. Table 9 presents observed counts (O) and observed/expected ratios (O/E) for $\#C_1V(N)C_2$ sequences in Malto, in which C_1 and C_2 are coronal plosives and N is an (optional) intervening homorganic nasal. Expected values are omitted to avoid cluttering the table. They can be found in Appendix B. Plosives are classified according to place of articulation (dental or retroflex) but laryngeal distinctions are ignored. The retroflex sonorant /ɽ/ is included for comparison with the plosives. It does not occur word-initially in Malto and is therefore restricted to C_2 position.

Table 9 Malto coronal plosives and /ɽ/ in $\#C_1V(N)C_2$ sequences ($n = 137$)

C_1\ C_2		t, d	ʈ, ɖ	ɽ
t, d	O	30	0	37
	O/E	2.04	0.00	1.61
ʈ, ɖ	O	0	60	10
	O/E	0.00	1.96	0.42

Table 9 should be read as follows. The vertical axis represents C_1 and the horizontal axis represents C_2. Observed counts (O) and observed/expected ratios (O/E) are listed for each possible C_1-C_2 configuration. Thus, there are 30 observed instances of a dental plosive in C_1 position followed by another dental plosive in C_2 position (t/d…t/d). The O/E ratio for this combination is 2.04, indicating that it occurs approximately two times more frequently than expected under random co-occurrence of consonants in the data set. Similarly, there are 60 observed instances of a retroflex plosive in C_1 position followed by another retroflex plosive in C_2 position (ʈ/ɖ…ʈ/ɖ). The O/E ratio for this combination is 1.96, indicating that it too occurs about two times more frequently than expected. In contrast to this, sequences consisting of a dental plosive followed by a retroflex plosive (t/d…ʈ/ɖ), or a retroflex plosive followed by a dental plosive (ʈ/ɖ…t/d), are both categorically absent (O = 0, O/E = 0.00). Thus, the data in Table 9 confirm Mahapatra's (1979) observation and reveal that the prohibition on co-occurring dental and retroflex plosives is categorical in Malto. Disharmonic T-Ṭ and Ṭ-T sequences are categorically absent while Ṭ-Ṭ sequences that agree in retroflexion and T-T sequences that agree in non-retroflexion are both over-attested.

Dental and retroflex plosives are more or less unconstrained with respect to all other consonants, coronal and non-coronal, including the retroflex sonorant /ɽ/. The retroflex sonorant does not participate in harmony either as a trigger or target of assimilation. In fact, the numbers in Table 9 suggest that disharmonic T-Ṛ sequences (O/E 1.61) are preferred over harmonic Ṭ-Ṛ sequences (O/E 0.42). The prohibition against T-Ṭ sequences combined with the well-formedness of T-Ṛ sequences is indicative of a similarity effect: retroflex consonant harmony holds only between highly similar coronal obstruents, namely dental and retroflex plosives, but not between plosives and sonorants. Examples illustrating the co-occurrence of dental and retroflex plosives with other coronal consonants are provided in (6).

(6) Co-occurrence of dental and retroflex plosives with other coronals in Malto

a. No harmony with palatal affricates[6]

ʧat-	‘to drip’	ʧuʈ-	‘to throw’
ʧaːd-	‘to select’	ʧeɖ-	‘to carry on shoulders’
ʤaːti	‘very tight’	ʤuʈi	‘a stone sling’
ʤuda	‘separated’	ʤuɖa	‘shade, shadow’

[6] Both T-Č (O = 4) and Ṭ-Č (O = 0) have low observed counts, while Č-T (O = 40) and Č-Ṭ (O = 49) are quite frequent. This surface pattern bears a resemblance to that of Pengo (cf. §3.1.2 below) in which *T-Č is avoided in favour of Č-Č but Č-T is preserved. Unlike Pengo, however, there appears to be no evidence from alternations or historical-comparative data to support directional palatal harmony in Malto. Most likely, the Malto numbers reflect a simple preference for palatals in initial over non-initial positions.

b. No harmony with coronal sibilants or nasals

teːs-	‘to fence’	ʈeski	‘proud’
setʁ-	‘to jump’	saːʈj-	‘to paste’
teːni	‘honey bee’	ɖaːni	‘wife, mistress’
nuːd-	‘to straighten’	nuɖ-	‘to hide’

c. No harmony with coronal liquids /l/ and /r/

teːl-	‘to shift’	ɖeːle	‘bulge’
lata	‘chance’	laʈa	‘gum resin’
toːrj-	‘to do effectively’	ʈuːrj-	‘to shave the head’
raːti	‘night’	roːɖo	‘cork tree’

d. No harmony with the retroflex flap /ɽ/

taɽe	‘grinding stone’	ʈeːɽa	‘dried seeds’
daɽe	‘animal for sacrifice’	ɖaɽa	‘tuber’

e. No harmony between coronal sonorants and /ɽ/

niɽg-	‘to powder, to grind’	noːɽ	‘to wash’
laɽar	‘to move, to shake’	loɽa	‘a stone to grind spices’
raːɽi	‘queen’	raːɽe	‘enemy’

(7) Co-occurrence of coronal and non-coronal plosives in Malto

topa	‘bunch’	ʈopa	‘drop of liquid’
deka	‘let’s go!’	ɖika	‘full, over-burdened’
gatj-	‘to stir, to mix’	gaʈj	‘to thread’
bed	‘to search’	beɖa	‘father’s father’

The examples in (6)(a–d) demonstrate that the coronal consonants /ʧ, ʤ, s, n, l, r/ and the retroflex sonorant /ɽ/ co-occur with both dental and retroflex plosives. Those in (6)(e) demonstrate that /ɽ/ also co-occurs freely with non-retroflex coronal sonorants such as /n, l/ and /r/. Finally, the examples in (7) demonstrate that dental and retroflex plosives co-occur freely with labial and dorsal consonants. Thus, apart from dental plosives, there is no indication that any other consonants serve as targets of retroflex assimilation; and apart from retroflex plosives, there is no indication that any other consonants serve as triggers of retroflex assimilation, including /ɽ/.[7]

There is clear evidence that retroflex consonant harmony is regressive in Malto. Disharmonic T-Ṭ configurations were frequent in Proto-Dravidian. Their absence in Malto, combined with the higher-than-expected frequency of Ṭ-Ṭ configurations, points to a process of regressive assimilation: initial dental plosives have become retroflex in Malto whenever they were followed by a retroflex plosive within the same root (i.e., T-Ṭ → Ṭ-Ṭ). This is evident when Malto roots are compared with cognates in other Dravidian languages. For instance, compare Malto /ʈuɖ-/ 'to smear' with Telugu /tuɖ(u)ʧu/ 'to wipe, rub' and Kannada /toɖe/ 'to smear'. Further examples are cited in §3.1.4, below, where the historical-comparative evidence of regressive retroflex consonant harmony in Dravidian is discussed in greater detail.

---

[7] The co-occurrence restriction on dentals and retroflexes might extend to dental /ð/. /ð/ does not occur word-initially and the overall count for /ð/ in C_2 position is low making O/E values for C_1…ð sequences unreliable. Nevertheless, observed counts suggest that /ð/ is categorically absent after retroflexes (O = 0) while it appears to be unrestricted after other coronal obstruents including dental plosives (O = 6) and palatal affricates (O = 5). It is not clear whether the absence of Ṭ…ð sequences is principled or whether it stems from the historic prohibition on word-initial retroflexes combined with the overall low frequency of /ð/.

While consonant harmony in Malto clearly involves regressive retroflex assimilation, it is not clear that it is limited to regressive retroflex assimilation. On the one hand, the absence of Ṭ-T sequences in Malto is not surprising; such sequences were also absent in Proto-Dravidian owing to the phonotactic restriction on word-initial retroflexes. On the other hand, the absence of Ṭ-T sequences in Malto is somewhat unexpected, given that the language has developed word-initial retroflex segments independent of retroflex harmony, and that Ṭ-P and Ṭ-K configurations are well attested (where P and K represent labial and dorsal plosives, respectively). The absence of Ṭ-T configurations could be attributed to either progressive retroflex assimilation (Ṭ-T → Ṭ-Ṭ), or regressive dental assimilation (Ṭ-T → T-T). However, there is no evidence for either of these sound changes. The most we can say is that the language never had Ṭ-T configurations historically, and never developed them. Beyond this, everything is speculation.

In sum, Malto exhibits a pattern of retroflex consonant harmony with the following properties: (i) it is root-internal; (ii) it is regressive (or possibly bidirectional); (iii) it holds only between pairs of coronal plosives, but not between plosives and sonorants; and (iv) it does not exhibit any known blocking effects.

3.1.2 Pengo (South-Central Dravidian)

Pengo is a South-Central Dravidian language spoken in the state of Orissa, India. It exhibits two patterns of consonant harmony: (i) a pattern of retroflex consonant harmony like that in Malto (§3.1.2.1); and (ii) a pattern of palatal consonant harmony (§3.1.2.2). Palatal harmony is independent of retroflex harmony in Pengo and does not involve retroflex consonants in any way. Nevertheless, it is included in the following discussion because it involves the long-distance assimilation of a coronal place feature that exhibits many parallels with the pattern of retroflex consonant harmony.

The consonant phonemes of Pengo are listed in (8).

(8) Consonant phonemes of Pengo (Burrow & Bhattacharya, 1970)

LAB	DEN/ALV	RET	PAL	VEL	GLOT
p	t	ʈ	ʧ	k	
b	d	ɖ	ʤ	g	
	s				
	z				h
m	n	ɳ		ŋ	
	l				
	r	ɽ			
w			j		

As shown in (8), Pengo distinguishes dental and retroflex consonants across three manners of articulation: plosives, nasals and rhotics. Like most South Asian languages, it also has a series of palatal stops that are realized phonetically as affricates.

3.1.2.1 Retroflex consonant harmony in Pengo

Burrow & Bhattacharya (1963) identify Pengo as one of several South-Central Dravidian languages in which "an initial dental is usually assimilated to a following cerebral [=retroflex]" (p. 240). A statistical analysis of the vocabulary list in Burrow & Bhattacharya's (1970) Pengo grammar confirms that retroflex consonant harmony is extensive in the language and exhibits the same typological properties found in Malto. Table 10 lists observed counts and O/E ratios for all word-initial $C_1V(N)C_2$ sequences in which C_1 and C_2 are both coronal plosives or retroflex sonorants, and N is a homorganic nasal.

Table 10 Pengo coronal plosives and retroflex sonorants in $\#C_1V(N)C_2$ sequences (n = 77)

C_1\ C_2		t, d	ʈ, ɖ	ɳ, ɽ
t, d	O	15	1	14
	O/E	2.57	0.08	1.20
ʈ, ɖ	O	0	31	16
	O/E	0.00	1.59	0.87

The data in Table 10 confirm that Pengo exhibits a pattern of retroflex consonant harmony like that of Malto. Disharmonic Ṭ-T and T-Ṭ sequences are either categorically absent or nearly so (O/E = 0.00 and 0.08, respectively) while Ṭ-Ṭ sequences that agree in retroflexion and T-T sequences that agree in non-retroflexion are both over-attested (O/E = 1.59 and 2.57, respectively). The single exception is the result of variation between disharmonic and harmonic forms of the same root (i.e., /tuːʈ- ~ ʈuːʈ-/ 'to scatter'). Representative examples are listed in (9). The examples in (9)(a) demonstrate the co-occurrence of two dental plosives. Those in (9)(b) demonstrate the co-occurrence of two retroflex plosives. Cognates from South Dravidian Tamil (Ta.) are provided in (9)(b) for the sake of comparison.

(9) Retroflex consonant harmony in Pengo

a. tuːt- 'to crouch down'

tiːd-ba 'to be cured' (with intensive suffix -ba)

daːda 'elder brother'

tandra 'necessity, compulsion'

b.	ʈu:ʈ- ~ tu:ʈ-	‘to scatter’	cf. Ta. tu:t̪t̪u	‘to scatter’
	ʈa:ʈi	‘mat’	cf. Ta. taʈʈi	‘screen’
	ʈoʈa	‘mango grove’	cf. Ta. to:ʈʈam	‘orchard’
	ʈoɖ-	‘to bow down (ears of paddy)’	cf. Ta. toɻu	‘to worship’[8]
	ʈi:nɖ-	‘to sharpen’	cf. Ta. ti:ʈʈu	‘to sharpen’

The Tamil cognates in (9)(b) represent phonologically conservative forms in which harmony has not applied. They demonstrate that harmonic Ṭ-Ṭ sequences in Pengo correspond to disharmonic T-Ṭ sequences with initial dental plosives elsewhere in Dravidian. Thus, at the very least, retroflex assimilation has applied regressively in Pengo (i.e., T-Ṭ → Ṭ-Ṭ). However, as in the case of Malto, retroflex plosives can occur word-initially before labial and dorsal plosives, as illustrated in (10). In light of this fact, the absence of Ṭ-T configurations is somewhat unexpected and could be taken to indicate progressive retroflex assimilation or regressive dental assimilation, although there is no clear evidence of either.

(10) Pengo: initial retroflex plosives before non-coronal consonants

ʈaku	‘stone of mango’	ɖu:ki	‘tree trunk, log’
ʈopa	‘red ant’	ɖa:bi	‘kneecap’

As in the case of Malto, consonant harmony does not apply across intervening morpheme boundaries in Pengo. This is evident from the examples in (11).

---

[8] Alternatively, Pengo /ʈoɖ-/ ‘to bow down (ears of paddy)’ might be cognate with Tamil /ta:ɻ-/ ‘to fall low’ and other items listed under etymological group 3178 in Burrow and Emeneau (1984).

(11) Pengo: no harmony across an intervening morpheme boundary

a. te-beʈ ‘then, therefore’[9]

te-benɖ-he ‘at that time’

b. ʈuj-t- past stem of ʈuj- ‘to suck (blood)’

ɖuː-t- past stem of ɖuː- ‘to touch’

All other coronals, including the retroflex sonorants /ɳ/ and /ɽ/, do not participate in the pattern of harmony either as triggers or targets of assimilation. The numbers in Table 10 suggest that disharmonic T-Ṛ sequences (O/E 1.20) are slightly preferred over harmonic Ṭ-Ṛ sequences (O/E 0.87), though both sequences occur relatively close to their expected frequencies. Examples are listed in (12) and (13).

(12) Pengo: no harmony with retroflex /ɳ/

a. taɳki ‘fold of skin hanging down from cock's neck’

henki duːɳa ‘hole in the ground for pounding’

b. ʈaːɳi ‘a quarrelsome person’

ʈuːɳ- ‘to slaughter, to sacrifice’

c. naːɳi ‘fire’

noːɳ- ‘to spin’

---

[9] Burrow & Bhattacharya’s (1970) vocabulary list includes /tebeʈ/ ‘then, therefore’ and /tebenɖ-he/ ‘at that time’, both with initial t-, along with /ebeʈtaŋ/ ‘from there, after that’ and /ebenɖ/ ‘at that time’, both without initial t-. They decompose the latter forms (without initial t-) into the demonstrative base /e-/ and a second element /benɖ- ~ beʈ-/ (p. 58) (/-taŋ/ in /e-beʈ-taŋ/ is an ablative suffix). I assume that this morphological analysis can be extended to the forms with initial t- although Burrow & Bhattacharya do not discuss these forms directly and the source of the initial t- remains unclear.

(13) Pengo: no harmony with retroflex /ɽ/

a. taɽi ‘mother’

toɽndel ‘sister’

b. ʈaːɽu ‘top of head’

ʈuːɽaː ‘to assemble’

Retroflex /ɳ/ co-occurs with both dental plosives, as shown in (12)(a), and retroflex plosives, as shown in (12)(b). The examples in (12)(c) demonstrate that retroflex /ɳ/ does not trigger harmony in initial dental nasals. Retroflex nasals do not occur word-initially in Pengo. Similarly, retroflex /ɽ/ co-occurs with both dental plosives, as shown in (13)(a), and retroflex plosives, as shown in (13)(b).[10] Pengo is unusual in exhibiting a few examples of word-initial /ɽ-/ but overall counts for /ɽ/ in that position are so low as to make statistical generalizations unreliable. For this reason they are not included in Table 10. Nevertheless, the few examples that do occur are all disharmonic Ṛ-T sequences (e.g., /ɽat-/ ‘to cause to weep’).

In sum, Pengo exhibits a pattern of retroflex consonant harmony with the same typological properties as that of Malto: (i) it is root-internal; (ii) it is regressive (or possibly bidirectional); (iii) it holds only between two coronal plosives, but not between plosives and sonorants; and (iv) it exhibits no known blocking effects.

3.1.2.2 Palatal consonant harmony in Pengo

Pengo also exhibits a pattern of regressive “palatal” harmony. Burrow & Bhattacharya (1970) report that initial dental stops are realized as palatal when followed by another palatal within

---

[10] There are no examples of /r…ɽ/ sequences in Burrow & Bhattacharya’s (1970) data.

the same root (i.e., T-Č → Č-Č). Some dialects of Pengo preserve the original dental with the result that dialectal variation between harmonic Č-Č and disharmonic T-Č can be found. Examples are provided in (14). Variation is not reported for the last example in (14) but disharmonic cognates are attested elsewhere in the South-Central group, as indicated by the cross-reference to Gondi (Go.).

(14) Palatal consonant harmony in Pengo (Burrow & Bhattacharya, 1970, p. 9)

ʧiʧ- ~ tiʧ-	past stem of tin- ‘to eat’[11]		
ʧinʤeŋ ~ tinʤeŋ	infinitive of tin- ‘to eat’		
ʧaːnʤ- ~ taːnʤ-	‘to weave (garland)’		
ʧonʤ- ~ tonʤ-	‘to appear’		
ʧoːʧ- ~ toːʧ-	‘to show’		
ʤoʧ-	‘to carry on the head’	cf. Go. toːʧ-	‘to carry on head’

Dialects that enforce palatal harmony prohibit *T-Č sequences, but all dialects (including those with palatal harmony) allow Č-T sequences, as shown in (15).

---

[11] Examples such as /ʧiʧ- ~ tiʧ-/ and /ʧinʤeŋ ~ tinʤeŋ/ are described as past and infinitive stems (respectively) of the root /tin-/ ‘to eat’. This might suggest that such forms are morphologically complex and that harmony applies across morpheme boundaries. However, the past and infinitive suffixes are /-t-/ and /-deŋ/ (respectively) with initial dentals, not palatals. These suffixes have allomorphs with initial palatals only as a result of local progressive assimilation following roots with final palatals (e.g., past and infinitive forms of /kiʧ-/ ‘to pinch’ are /kiʧʧ-/ (< kiʧ-t-) and /kiʤʤeŋ/ (< kiʧ-deŋ)). A small class of verbs ending in /n/, that includes /tin-/ ‘to eat’, forms its past and infinitive stems with palatal stops (cf. /maʧ-/ and /manʤeŋ/, the past and infinitive forms of /man-/ ‘to be’). Burrow & Bhattacharya explain that this is the result of an “ancient sandhi” (1970, pp. 97, cf. 66). The source of this sandhi is no longer transparent. Thus, stem forms such as /ʧiʧ- ~ tiʧ-/ and /ʧinʤeŋ ~ tinʤeŋ/ may not be considered morphologically complex from a synchronic point of view. If this is so, then the harmony holds only over morpheme-internal sequences.

(15) Pengo: No harmony in Č-T sequences (cf. Hansson 2001, p. 85)

ʧeta man-	'to be awake'	~ *ʧeʧa man-	~ *teta man-
ʧinta ki-	'to think, to worry'	~ *ʧinʧa ki-	~ *tinta ki-
ʤuːt-	'to bring down, to put down'	~ *ʤuːʧ-	~ *duːt-
ʤunda	'spinning top'	~ *ʤunʤa	~ *dunda

There are two asymmetries to the pattern. First, there is a trigger-target asymmetry: dentals assimilate to palatals, but palatals do not assimilate to dentals (i.e., Č-T ↛ T-T). Second, as Hansson (2001; 2010) points out, the harmony is purely regressive; dentals assimilate to a following palatal but not to a preceding one (i.e., Č-T ↛ Č-Č). The same pattern of palatal harmony, showing the same dialectal variation and asymmetries, has also been reported in Kuvi, another Dravidian language of the South-Central group (Burrow & Bhattacharya, 1963, p. 233).[12] Representative examples are shown in (16) and (17) (with data from Burrow & Bhattacharya, 1963; and Israel, 1979).

(16) Palatal consonant harmony in Kuvi

ʧaːʧ-	~ taːʧ-	'to sew'
ʧuʧ-	~ tuʧ-	'to block up'
ʧoːnʤ-	~ toːnʤ-	'to appear'
ʤuːʧ-	~ duʧʧ-	'to carry on head'
ʧoʤo	~ toʤo	'floor'

[12] Languages outside of South Asia with similar types of palatal harmony showing the same directional and trigger-target asymmetries include Bolivian Aymara (Aymaran) and Kera (West Chadic) (Hansson, 2001; 2010).

(17) Kuvi: No harmony in Č-T sequences

ʤaːti	'caste'	~ *ʤaːʧi	~ *daːti
ʤiːtomi	'fees, salary'	~ *ʤiːʧomi	~ *diːtomi
ʤeta	'grinding stone'	~ *ʤeʧa	~ *deta
ʤonto	'animal'	~ *ʤonʧo	~ *donto

In sum, Pengo exhibits both retroflex and palatal consonant harmony. The two patterns share many typological properties including: (i) they are both root-internal; (ii) they are both regressive (although retroflex harmony is *possibly* bidirectional); (iii) they are both triggered by postalveolar coronal stops and they both target dental plosives; and (iv) neither exhibits any known blocking effects.

3.1.3 The scope of retroflex consonant harmony in Dravidian

The preceding sections reviewed evidence indicating that Malto, a North Dravidian language, and Pengo, a South-Central language, both exhibit the same pattern of retroflex consonant harmony. A survey of Dravidian reveals that these two languages are not exceptional; they are representative of their respective subgroups. Retroflex consonant harmony of the kind exhibited by Malto and Pengo is characteristic of most Dravidian languages in the North and South-Central subgroups, and extends to at least one language of the Central group: Parji.

North Dravidian is the smallest subgroup of the family, consisting of only three major languages: Malto, Kurux and Brahui. Of these, Brahui is geographically isolated in western Pakistan and does not exhibit evidence of consonant harmony. However, retroflex consonant harmony has been reported in Kurux (Pfeiffer, 1972, pp. 83, 153), which is spoken in eastern India along with Malto. In addition, Burrow & Bhattacharya (1963) report retroflex consonant harmony in Kuvi, a language of the South-Central group, which is also spoken in eastern India in close proximity to Pengo. Significantly, they point out that the pattern of retroflex consonant harmony in Kuvi is characteristic of "most of the Dravidian languages of this area" (1963, p. 240). The languages they name include the South-Central languages, Kuvi, Kui, Pengo and Konda, and the Central Dravidian languages, Parji and Gadaba. At least one other South-Central language, Gondi, also shows a tendency toward retroflex harmony, though no mention of it was found in the literature.[13]

A survey of data from all these languages confirms the presence of retroflex consonant harmony in each one, with the exception of Central Dravidian Gadaba. Table 11 shows the co-occurrence of coronal plosives and retroflex sonorants in $\#C_1V(N)C_2$ sequences from eight Dravidian languages with retroflex consonant harmony. For ease of readability the results are presented schematically following a convention introduced by Pozdniakov & Segerer (2007). Rather than presenting observed, expected and O/E values, Pozdniakov & Segerer measure the discrepancy between observed and expected values and express it as a percentage, whether positive or negative. 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%.[14] For ease of readability, the results are then presented as follows: (i) a discrepancy whose absolute value is

---

[13] Limited data for Manda in Burrow & Emeneau (1984) suggests that it too belongs to the group of South-Central languages with retroflex consonant harmony, though the data is insufficient for a reliable analysis.

[14] The formula employed by Pozdniakov & Sergerer (2007) is: $100 \times ((O - E) \div E)$. They suggest that their method has an advantage over the χ^2 test because it is more readable and because it preserves the direction of deviation from the norm (expressed in terms of $+$ or $-$), which χ^2 does not (2007, p. 314, footnote 6).

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.[15] To further aid readability, cells with under-attested values are shaded grey. Wherever an expected value is lower than 5.0, O/E ratios and O/E discrepancies are potentially exaggerated. Here and elsewhere parentheses are used to mark such values. Observed and expected values for each language can be found in Appendix B.

Table 11 Coronal plosives and retroflex sonorants in #$C_1V(N)C_2$ sequences in eight Dravidian languages with retroflex consonant harmony[16]

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

	T	Ṭ	Ṛ
T	++	−−	
Ṭ	−−	++	

Malto (Mahapatra, 1987; n = 137)

	T	Ṭ	Ṛ
T	++	−−	++
Ṭ	−−	++	−−

[15] In Pozdniakov & Sergerer (2007) the thresholds are ±15% and ±30% instead of ±25% and ±50%, respectively. However, the lower thresholds employed by Pozdniakov & Segerer fail to capture an important distinction in the current data: the distinction between a language with categorical or near-categorical avoidance of a given C_1-C_2 combination (e.g., −95%) and one with only a relative dispreference for the combination (e.g., −35%). Using the lower thresholds of 15% and 30%, both of these would be represented as "−−". For this reason I have increased the thresholds to 25% and 50%. This reserves each category only for those combinations that show a much higher degree of over- or under-attestedness.

[16] For each language in Table 11, the class of retroflex sonorants represented by Ṛ includes the retroflex flap /ɽ/. In Kui, Kuvi, Pengo and Konda it also includes /ɳ/.

Kui (Winfield, 1929; Burrow & Bhattacharya 1961; Burrow & Emeneau, 1984; n = 42) [17]

	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	− −	+ +	− −

Several observations can be made based on the results in Table 11. To begin with, all of the languages in Table 11 exhibit a pattern of retroflex consonant harmony between coronal plosives but not between plosives and sonorants. For example, in Kurux, Ṭ-T and T-Ṭ sequences, 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 sequences that agree in retroflexion, and T-T sequences that agree in non-retroflexion, are both represented by double "+ +" signs, indicating that they are over-attested by more than +50%.

---

[17] In the case of Kui, no individual source contained enough data for a reliable statistical analysis so data was combined from three sources (i.e., Winfield, 1929; Burrow & Bhattacharya, 1961; and Burrow & Emeneau, 1984). Where the same word was listed in more than one source it was counted only once. In an effort to further increase counts a few examples of #C_1rV(N)C_2 were included (e.g., ɖɽeːnɖu 'tail', traːɖa 'liver', etc.). In spite of these things, overall counts for this language remain low. However, the trend that they reveal is consistent with that of the other South-Central Dravidian languages that exhibit retroflex consonant harmony.

Thus, sequences of two coronal plosives that agree in place of articulation are overwhelmingly preferred over those that disagree in place. The same trend holds for every other language in the table. In each case, disharmonic Ṭ-T and T-Ṭ sequences are under-attested (in most cases categorically absent or nearly so) while harmonic Ṭ-Ṭ and T-T sequences are over-attested.

A very different pattern holds when retroflex sonorants are involved. In most cases, disharmonic T-Ṛ sequences are preferred over harmonic Ṭ-Ṛ sequences. This is indicated by "+" signs (single or double) attributed to T-Ṛ sequences in most of the languages, and by "–" signs (single or double) attributed to Ṭ-Ṛ sequences. This pattern is precisely the opposite of that exhibited for pairs of plosives.

In the case of Kurux and Pengo, disharmonic T-Ṛ sequences and harmonic Ṭ-Ṛ sequences are represented by empty cells. This indicates that observed counts for these sequences fall within ±25% of their expected frequencies. Each one occurs more-or-less as expected and one is not strongly preferred over the other. Thus, retroflex sonorants co-occur with coronal plosives either as expected, or else disharmonic T-Ṛ sequences are preferred. No language shows a preference for Ṭ-Ṛ sequences. Thus, retroflex consonant harmony holds between coronal plosives but not between plosives and sonorants.

Secondly, while Gondi clearly exhibits evidence of retroflex consonant harmony, it exhibits a much weaker pattern of harmony than any of the other languages. It is the only language in Table 11 in which disharmonic T-Ṭ sequences are represented by a single "–" sign (as opposed to a double "– –" sign). This indicates that these sequences are not under-attested to the same degree that they are in other languages with retroflex consonant harmony. Gondi appears to preserve a large number of disharmonic T-Ṭ sequences alongside the more innovative Ṭ-Ṭ sequences. The mixed pattern exhibited in Gondi may be the result of dialectal variation with inter-borrowing between harmonic and disharmonic varieties. Many words in Penny et. al.'s (2005) Gondi dictionary are listed as having both harmonic and disharmonic variants. The examples in (18) are listed in the dictionary under both harmonic and disharmonic spellings with cross-referencing between the two.[18]

(18) Variable retroflex consonant harmony in Gondi (data from Penny et. al., 2005)

taːɖvaː	~	ʈaːɖvaː	'chin'
toɖɖi	~	ʈoɖɖi	'beak, mouth'
teɳɖ-	~	ʈeɳɖ-	'to take out, remove, draw (water)'
daːʈ	~	ɖaːʈ	'much, many'
daɳɖaːri	~	ɖaɳɖaːri	'Gond ritual song and dance'
daːɳɖa	~	ɖaːɳɖa	'handle, shaft (of tool such as an axe, or digger)'
dʰeɳɖeː	~	ɖeːɳɖeː	'stalk'

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 & Bhattacharya, 1963). Moreover, retroflex consonant harmony does not apply to any language of the South Dravidian group. South Dravidian is by

[18] Data for the present study is drawn from a dictionary of the Southern (Adilabad) variety of Gondi compiled by Penny et. al. (2005), which in turn is "based on, but not limited to" the wordlists contained in Lincoln (1969) and Subrahmanyam (1968). A mix of harmonic and disharmonic forms can also be found in other sources on Gondi dialects. Cf. data in Burrow & Bhattacharya (1960) and Burrow & Emeneau (1984).

far the largest subgroup. It accounts for more than half of all Dravidian languages and includes three of the four major literary languages: Tamil, Malayalam and Kannada.

All Dravidian languages that lack retroflex consonant harmony exhibit a pattern like the one attributed to Proto-Dravidian in (2); retroflex consonants are absent word-initially (or very nearly so) and disharmonic T-Ṭ sequences are abundant. Table 12 shows observed counts for coronal plosives and retroflex sonorants in $\#C_1V(N)C_2$ sequences from three Dravidian languages that lack retroflex consonant harmony: Telugu (representing a disharmonic variety of South-Central Dravidian), Gadaba (representing Central Dravidian) and Tamil (representing South Dravidian).[19]

Table 12 Observed counts for coronal plosives and retroflex sonorants in $\#C_1V(N)C_2$ sequences from Dravidian languages that lack 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

[19] O/E ratios are not reported in Table 12 because they are misleading in cases where retroflex consonants are avoided altogether in word-initial position. See discussion in §1.4.

The observed counts in Table 12 support the conclusion that these languages do not exhibit retroflex consonant harmony. In each case, disharmonic T-Ṭ sequences have the highest observed count while harmonic Ṭ-Ṭ sequences are among the lowest.

Recall that Burrow & Bhattacharya (1963) list Gadaba among the Dravidian languages with retroflex consonant harmony. The results of the present survey do not bear this out. The single example cited by Burrow & Bhattacharya is /ʈeʈp-/ 'to raise, lift', which appears to be an exception. There are at least two varieties of Gadaba: Ollari and Koṇekor (a.k.a. Mudhili). The results in Table 12 are based on counts of Koṇekor Gadaba but harmony does not appear to hold for either variety. In Bhattacharya's (1957) Ollari Gadaba vocabulary, /ʈeʈp-/ 'to raise, lift' and /ʈoʈo vande/ 'thumb, big toe' are the only examples showing retroflex harmony. All other words containing retroflex plosives are disharmonic, including /toːʈp-/ 'to show', /toːnɖ/ 'kinsman', /tinɖ-/ 'to pull', etc. In Bhaskararao's (1980) Koṇekor Gadaba vocabulary all but one of the words containing retroflex plosives are disharmonic, including /teːʈp-/ 'to lift'. Other sources also list a mix of harmonic and disharmonic forms for Gadaba (Burrow & Bhattacharya, 1962; Burrow & Emeneau, 1984) but in each case disharmonic forms appear to be the norm and harmonic forms the exception.

In sum, retroflex consonant harmony is characteristic of most languages of the North Dravidian group, including Malto and Kurux (but not Brahui); most languages of the South-Central Dravidian group, including Kui, Kuvi, Pengo, Konda and (to a lesser extent) Gondi (but not Telugu); and at least one language of the Central Dravidian group: Parji. It does not appear to affect any other Central Dravidian language, including Gadaba, nor does it affect any language of the South Dravidian group, which constitutes the majority of Dravidian languages.

3.1.4 Historical-comparative evidence of consonant harmony

Historical-comparative data supports the conclusion that retroflex consonant harmony has applied in most of the North and South-Central Dravidian languages and in Central Dravidian Parji. Harmonic Ṭ-Ṭ forms in these languages correspond to disharmonic T-Ṭ forms in South Dravidian languages, such as Tamil and Malayalam, or in those South-Central or Central Dravidian languages where harmony has not applied, such as Telugu and Gadaba. Representative examples from North, South-Central and Central Dravidian are listed in (19), (20) and (21), respectively. In each data set, harmonic word forms from the respective language group are identified with disharmonic cognates from Telugu (representing a conservative “disharmonic” variety of South-Central Dravidian) and Tamil (representing South Dravidian). Tamil is the oldest literary Dravidian language and is generally regarded as one of the most phonologically conservative (i.e., closest to Proto-Dravidian). In the few instances where a cognate from Tamil was not available, a cognate from Malayalam has been supplied instead (indicated by the abbreviation “Ma.”). Reference numbers in the right-most column of each data set refer to etymological groups in Burrow & Emeneau’s (1984) revised *Dravidian Etymological Dictionary* (DEDR).

(19) Consonant harmony in North Dravidian: Kurux and Malto

	Kurux	Malto	Telugu	Tamil	DEDR
'to grope'	–	ʈoʈr-	taɖavu	taʈavu	3025[20]
'to hinder'	ʈaɳɖna:	–	taʈa:jintʃu	taʈu, taɳʈu	3031
'side'	ʈo:ʈ	–	taʈʈu	taʈʈu	3040
'strength'	ʈinɖi:	–	daɳɖi (?)	tiɳ, tiɳʈi	3222
'to smear, wipe'	ʈu:ɽna:, ʈuɖɖjas	ʈuɖ	tuɖutʃu	tuʈai	3301
'big(-bellied)'	ɖonɖa:	ɖuɖwa	doɖɖa, doɖɖu	toʈʈa	3491

(20) Consonant harmony in South-Central Dravidian: Kuvi, Konda & Gondi

	Kuvi	Konda	Gondi	Telugu	Tamil	DEDR
'many, thick'	–	ɖaʈam	ɖa:ʈ, da:ʈ	daʈʈamu	taʈam	3020
'upper arm'	ɖaɳɖe	ɖaɳɖa	ɖanɖ, danɖ	daɳɖa	taɳʈa (Ma.)	3048
'kind of tree'	ʈa:ɳɖi	ʈa:ɳɖi	–	tã:ɖi, ta:ɳɖra	ta:n̺t̺i	3198
'stout stick'	ɖuɖu	ɖuɖu	ɖuɖu	duɖɖu	tuʈʈu-kkaʈʈai	3304
'backyard'	ɖoɖi	ɖoɖi	ɖoɖɖi, ɖiɖɖi	doɖɖi	toʈʈi	3485
'beehive'	ʈaʈʈi, ʈaʈʈo	te:ne-ʈaʈa	ʈeʈʈe, teʈʈe	te:ne-teʈʈu	ti:n-toʈai	3490
'big'	ɖoʈa	–	–	doɖɖa, doɖɖu	toʈʈa	3491
'creeper'	–	ɖoɳɖa	ʈonɖa, ʈonɖri	doɳɖa	toɳʈai	3499
'garden'	ʈo:ʈa	ʈo:ɳʈa	–	tõ:ʈa	to:ʈʈam	3549
'companion'	to:ɽu	to:ɽu	-to:ɽo	to:ɖu	to:ɹan̺	3563

[20] Alternatively, 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 (21).

(21) Consonant harmony in Central Dravidian: Parji

	Parji	Gadaba	Telugu	Tamil	DEDR
‘upper arm’	ɖanɖa	–	daɳɖa	taɳʈa (Ma.)	3048
‘to pull’	ʈanɖ-	tinɖ-	–	–	3052
‘to touch’	ʈoɖ-, ʈoʈʈ-	–	toʈʈu	toʈu, toʈʈ-	3480
‘creeper’	ʈunɖa	–	doɳɖa	toɳʈai	3499
‘labourer’	ʈo:ʈal, to:ʈa	–	to:ʈi	to:ʈʈi	3546
‘rake/hook’	ʈo:ʈal	–	dõ:ʈi	to:ʈʈi	3547
‘to draw water’	ʈõ:ɖ-	to:ɳ-, to:nɖ-	tõ:ɖu	to:ɳʈu	3549
‘to show’	ʈo:ʈip-	to:ʈp-	–	to:t̪t̪u (to:t̪t̪i-)	3566

A few observations can be offered concerning the correspondences in (19), (20) and (21). To begin with, all of the historical-comparative evidence points toward regressive retroflex assimilation (T-Ṭ → Ṭ-Ṭ). Synchronically, the coronal co-occurrence patterns are ambiguous with respect to direction. On the one hand, the general absence of Ṭ-T configurations is expected, given the historical prohibition on word-initial apicals in Dravidian. On the other hand, the absence of Ṭ-T configurations is unexpected, given that most of the languages with consonant harmony have developed Ṭ-P and Ṭ-K configurations through the introduction of word-initial retroflex stops independent of harmony. The absence of Ṭ-T configurations might be attributed to progressive retroflex assimilation (Ṭ-T → Ṭ-Ṭ) or regressive dental assimilation (Ṭ-T → T-T). However, there is no historical-comparative evidence supporting either of these assimilation patterns in Dravidian. The most we can say is that Dravidian languages lacked Ṭ-T configurations historically and failed to develop them. This may reflect a principled avoidance of Ṭ-T configurations, but we cannot derive any conclusions about directionality from this supposition. Moreover, palatal consonant harmony in Pengo and Kuvi is strictly regressive despite the fact that the conditions for progressive palatal assimilation or regressive dental assimilation are met (see §3.1.2.2). Thus, it remains possible that retroflex consonant harmony in Dravidian is also strictly regressive.

Secondly, the triggers and targets of harmony are always plosives. With few exceptions, most cases of consonant harmony can be traced to forms in which the triggering plosive in C_2 position was either geminate or post-nasal. Single intervocalic plosives are often subject to lenition in Dravidian and other South Asian languages. This typically takes the form of flapping in the case of apical plosives, whether apico-alveolar or retroflex, and spirantization in the case of others. However, geminate and nasal-stop clusters are precisely the environments in which non-initial plosives consistently resist lenition. Thus, geminate and post-nasal retroflex plosives are more likely to trigger assimilation in initial dental plosives because they retain a similarity to them in terms of manner. Notice that the retroflex approximant /ɻ/ in roots such as Tamil /toːɻan̠/ corresponds to a retroflex flap /ɽ/ in Kuvi and Konda /toːɽu/ 'companion' and Gondi /-toːɽo/ (in (20)). The retroflex sonorants fail to trigger harmony in these cases.

Thirdly, harmony has been triggered by former apico-alveolar plosives in cases where they have developed into retroflex plosives. The phoneme represented by /t̠/ in Tamil /taːn̠t̠i/ (in (20)) and Tamil /toːt̠t̠u, toːt̠t̠i-/ (in (21)) is a voiceless apico-alveolar stop in gemination, a voiced apico-alveolar stop post-nasally and a voiced trill intervocalically. This phoneme was part of the Proto-Dravidian consonant system. It is preserved only in a few South Dravidian languages. Elsewhere it has merged variously with dentals or retroflexes. In cases where the old Dravidian alveolars have developed into retroflex plosives, they have triggered retroflex assimilation in any preceding dental plosives. This is evident in examples such as Kuvi/Konda /ʈaːɳɖi/ ‘kind of tree’ which corresponds to Tamil /taːn̪t̪i/ [taːn̪d̪i] (in (20)) and Parji /ʈoːʈip-/ ‘to show’ which corresponds to Tamil /toːt̪t̪u, toːt̪t̪i-/ (in (21)).

In sum, evidence from historical-comparative data supports the conclusion that the co-occurrence restriction on coronal plosives in North and South-Central Dravidian languages, and in Central Dravidian Parji, is the product of retroflex consonant harmony. It also confirms that retroflex consonant harmony in Dravidian is regressive and holds only between plosives.

3.1.5 Similarity effects: retroflex sonorants in Dravidian

The synchronic co-occurrence patterns summarized in Table 11 and the historical-comparative data reviewed in the preceding section both suggest that retroflex consonant harmony in Dravidian is sensitive to the similarity of participating segments in terms of their manner of articulation. In each of the cases reviewed, retroflex harmony holds only between coronal plosives. Harmony is not triggered by retroflex sonorants, whether /ɽ, ɳ/ or /ɭ/, despite the fact that each of the languages surveyed distinguishes one or more of these segments in its phonemic inventory. In every case Ṭ-Ṛ sequences occur at or below expected frequencies while disharmonic T-Ṛ sequences occur at or above expected frequencies. Thus, Dravidian languages tend to show a preference for disharmonic forms where retroflex sonorants are involved. However, some studies have suggested, either explicitly or implicitly, that retroflex sonorants have triggered regressive consonant harmony in word-initial plosives. These warrant some discussion.

Zvelebil (1970, p. 102) reports that some instances of initial retroflex /ʈ-/ or /ɖ-/ in Dravidian stem from Proto-Dravidian dental */t-/ when it was followed by a non-adjacent retroflex nasal */ɳ/. In support of this claim Zvelebil cites the example of Telugu /ɖakku/ 'to be obtained', which, according to his analysis, derives from Proto-Dravidian */taɳ-k-/. This derivation entails several intermediate stages of development. Minimially, these must include: (i) non-local regressive retroflex assimilation triggered by the nasal (i.e., */taɳ-k-/ > */ʈaɳ-k-/), and (ii) subsequent local assimilation of the retroflex nasal to the following velar plosive, which was originally part of a derivational suffix (i.e., */ʈaɳ-k-/ > */ʈaŋ-k-/ > */ʈak-k-/). With a few other minor modifications this would yield Telugu /ɖakku/.

There are several complications that raise doubts about Zvelebil's analysis. Granted, there is evidence to support the reconstruction of medial */ɳk/ (or */ɭk/) for some instances of medial /ŋk/ and /kk/ in the modern Dravidian languages. However, in all such cases it is possible to find cognates that have preserved a retroflex segment in that position (see examples cited by Zvelebil, 1970, p. 173). This is not the case for Zvelebil's reconstructed root */taɳ-k-/. All cognates of Telugu /ɖakku/ listed in Zvelebil (1970, p. 102) and in Burrow & Emeneau (1984, see DEDR 3014) contain velar nasals or plosives, with the exception of Malayalam which also includes some cognates with palatal nasals. Significantly, no language exhibits evidence of a retroflex segment corresponding to Zvelebil's reconstructed */ɳ/ in this root.[21] Moreover, Gwynn's (1991) Telugu dictionary does not record any instance of word-initial /ʈVɳ/ or /ɖVɳ/. It records only a single instance of initial /ʈVɭ/ (i.e., /ʈaːɭaːʈoːɭiːgaː/ ~ /ʈʰaːɭaːʈʰoːɭiːgaː/ 'lightheartedly'). By comparison, the word-initial sequences /tVɳ/, /dVɳ/, /tVɭ/ and /dVɭ/ are all

---

[21] Hamann (2003, p. 123) wrongly identifies the Tamil and Malayalam cognates of Telugu /ɖakku/ 'to be obtained' as /taɳku/ and /taɳːuka/, respectively, both with medial retroflex nasals. This appears to be a misinterpretation of Zvelebil's (1970) transcription. Zvelebil represents the nasals in question as /ṅ/, which denotes a velar nasal in the South Asian tradition (i.e., IPA [ŋ]).

well attested. Thus, there is nothing to indicate that retroflex sonorants have triggered retroflexion in word-initial dental plosives in Telugu, as suggested by Zvelebil, and we must assume that the initial retroflex plosive in /ɖakku/ has some other origin.

Other examples cited elsewhere in the Dravidian literature also appear to suggest that harmony has been triggered by retroflex sonorants, at least in some cases. Consider the examples in (22), the first two of which are cited in Burrow & Bhattacharya (1963, p. 240) and the last of which is cited in Pfeiffer (1972, pp. 63, 83).

(22) Apparent cases of retroflex consonant harmony triggered by sonorants

Kuvi	ʈuɳ-	‘to cut’	cf. Tamil	tuɳi	‘to be cut’
Kui	ʈuːɳu	‘log of wood’	cf. Tamil	tuːɳ	‘post’
Kurux	ʈõːɽ	‘to hook in’	cf. Tamil	toɭ, toːɭ	‘to perforate’

One possible interpretation of these data is to assume that consonant harmony has applied sporadically or in a gradient fashion between retroflex sonorants and dental plosives. However, the weight of evidence suggests that retroflex consonant harmony has not been triggered directly by the sonorants in these cases.

Variation between retroflex sonorants and retroflex plosives, or consonant clusters containing retroflex plosives, is relatively common, both within languages and between cognates in different languages. Some typical patterns of variation include: ɳʈ/ɳɖ ~ ɳ (e.g., Gadaba /taɳɖeka/ ~ /taɳaka/ ‘bamboo curtain’) and (ɳ)ɖ ~ (Ṽ)ɽ (e.g., Kurux /aɖ-/ ~ /aɽ-/ ‘to furnish with skin', /aɳɖ-/ ~ /ãːɽ-/ 'to spread throughout').[22] For most (if not all) examples that appear to show harmony triggered by retroflex sonorants it is possible to find cognates where the sonorant in question corresponds to a retroflex plosive or a cluster containing a retroflex plosive. Thus, in addition to Tamil /tuɳi/ 'to be cut' (DEDR 3305) we also find forms such as Tamil /tuɳʈi/ and Telugu /tuɳɖintʃu/, both meaning 'to cut' (DEDR 3310). If the word-initial retroflex plosive in Kuvi /ʈuɳ-/ 'to cut' is the product of harmony (as it probably is), then it is more plausible to assume that it developed from a form like *tuɳɖ- with a retroflex plosive in C_2 position. If that plosive was lost after triggering harmony it would leave behind the retroflex nasal in C_2 position, creating the appearance of harmony between the nasal and the initial plosive (i.e., *tuɳɖ- > *ʈuɳɖ- > ʈuɳ-).

A similar case can be made for most other forms that appear to exhibit harmony conditioned by retroflex sonorants. For example, Kurux /ʈõːɽ-/ 'to hook in' might not be cognate with Tamil /toɭ/ 'to perforate' (DEDR 3528, as suggested by Pfeiffer, 1972, p. 63) but rather with Tamil /toːʈʈi/ 'elephant hook or goad' (DEDR 3547). The nasalization of the vowel in Kurux suggests that it probably derives more directly from a form like that of Tulu /doːɳʈi/ 'long pole with hook to pluck fruit' (also DEDR 3547) with the nasal consonant absorbed by the vowel and subsequent lenition of the retroflex plosive in C_2 position (i.e., *toːɳɖ- > *ʈoːɳɖ- > *ʈõːɖ- > ʈõːɽ-). This development is supported by evidence from synchronic variation within Kurux itself. Most forms that appear to show harmony triggered by a retroflex sonorant actually have variants in which the sonorant corresponds to a plosive, as shown in (23).

---

[22] Sources for the examples cited here are: Bhaskararao, 1980, p. 13; Pfeiffer, 1972, p. 11. Cf. also the discussion in Burrow & Bhattacharya, 1953, pp. 6–7.

(23) Correspondences between -(Ṽ)ɽ and -(ɳ)ɖ in Kurux (data from Grignard (1986 [1924])

a. ʈãɽnaː ~ ʈaɳɖnaː 'to prevent, hinder'

ɖĩːɽ ~ ɖiːɳɖ 'a line'

ɖeɽẽː ~ ɖeɳɖeː '[??]'

b. ɖaɽkaː ~ ɖaɖkaː 'manger, trough'

ɖiɽʰ ~ ɖiɖʰ 'courage'

c. ʈuːɽnaː, ʈuɖɖjas 'to mark' cf. Malto ʈuːɖ 'to smear'

ʈiɽnaː, ʈiɖɖas 'to let fall' cf. Malto ʈiɖ- 'to vanquish'

A careful examination of the data reveals other significant correlations. Some languages have both harmonic and disharmonic word forms that are related historically to a common root. In these cases, the disharmonic form is inevitably the one in which the retroflex segment is realized as a sonorant, while the harmonic form is the one in which it is realized as a plosive. For example, Pengo has /toɽndel/ 'sister' but /ʈonɖen/ 'brother' (cf. Gadaba /toːnɖud/ 'sister', /toːnɖoːnɖ/ 'brother', DEDR 3563). Similarly, Pengo has /taɽi/ 'mother' (with retroflex sonorant and no harmony) while closely related Kuvi has /ʈaɖi/ 'mother' (with retroflex plosive and harmony), both of which are cognate with Tamil /taɭɭai/ 'mother' (DEDR 3136).

Not only do retroflex sonorants fail to trigger harmony, they also fail to participate as targets of harmony. This is evident from the fact that retroflex sonorants do not occur (or are extremely rare) in word-initial position in most Dravidian languages. Disharmonic R-Ṛ and R-Ṭ sequences are the norm while *Ṛ-Ṛ and *Ṛ-Ṭ sequences are generally unattested (cf. data for Malto in (6) and for Pengo in (12)). While retroflex plosives trigger harmony in other coronal plosives, they do not trigger harmony in sonorants. If they did, we would expect a dearth of R-

Ṭ sequences and an abundance of Ṛ-Ṭ sequences in their place. As it is, R-Ṭ and R-T are both well attested in every case, as are Ṭ-R and T-R.

In sum, the weight of evidence suggests that retroflex consonant harmony holds only between coronal plosives in Dravidian. Sonorants do not participate in harmony either as triggers or targets of assimilation. Harmony between plosives and sonorants is sporadic at best and even the few apparent cases that exist are doubtful. Most of these can be attributed to harmony between plosives with subsequent loss or lenition of the triggering plosive.

3.1.6 Laryngeal features

While retroflex consonant harmony is clearly sensitive to manner of articulation in Dravidian, laryngeal features play no role. Agreement for laryngeal features is neither a condition for retroflex assimilation nor a necessary output of it. This is evident in Table 13, which shows pairs of coronal plosives classified according to agreement for retroflexion and voicing in eight Dravidian languages that exhibit retroflex consonant harmony. In Table 13, plosive pairs that agree in retroflexion (Ṭ-Ṭ) are isolated from all others because they are the only ones that are clearly the products of retroflex consonant harmony. There is no evidence that T-T pairs agreeing in non-retroflexion are the product of consonant harmony. They represent the residue of harmony after disharmonic T-Ṭ sequences have become Ṭ-Ṭ through retroflex assimilation. Thus, pairs of plosives that agree in non-retroflexion (T-T) are classified together with those that disagree in retroflexion (T-Ṭ, Ṭ-T) under the label “Other”, which represents all pairs that are not the product of retroflex assimilation. Agreement for voicing in Table 13 includes agreement for both the presence and absence of voicing.

Table 13 Coronal plosives in $\#C_1V(N)C_2$ sequences in Dravidian languages classified according to agreement for retroflexion and voicing

Kurux (Grignard [1924] 1986; n = 190)

		Retroflexion	
Voicing		Agree	Other
Agree	O	77	71
	O/E	3.05	1.02
Disagree	O	19	23
	O/E	0.77	0.33

Malto (Mahapatra 1987; n = 90)

		Retroflexion	
Voicing		Agree	Other
Agree	O	41	24
	O/E	2.12	0.96
Disagree	O	19	6
	O/E	0.92	0.24

Kui (Winfield 1929, Burrow & Emeneau 1984; n = 31)

		Retroflexion	
Voicing		Agree	Other
Agree	O	16	3
	O/E	1.29	(0.91)
Disagree	O	9	3
	O/E	0.89	0.58

Pengo (Burrow & Bhattacharya 1970; n = 47)

		Retroflexion	
Voicing		Agree	Other
Agree	O	18	11
	O/E	1.82	0.84
Disagree	O	13	5
	O/E	1.16	0.39

Parji (Burrow & Bhattacharya 1953; n = 73)

		Retroflexion	
Voicing		Agree	Other
Agree	O	25	16
	O/E	2.23	0.75
Disagree	O	15	17
	O/E	1.27	0.59

Kuvi (Israel 1979; n = 57)

		Retroflexion	
Voicing		Agree	Other
Agree	O	26	7
	O/E	1.70	0.67
Disagree	O	15	9
	O/E	0.84	0.67

Gondi (Penny et. al. 2005; n = 70)

Voicing		Retroflexion: Agree	Retroflexion: Other
Agree	O	21	25
	O/E	1.90	1.01
Disagree	O	9	15
	O/E	0.94	0.61

Konda (Krishnamurti 1969; n = 21)

Voicing		Retroflexion: Agree	Retroflexion: Other
Agree	O	10	4
	O/E	1.57	0.81
Disagree	O	6	1
	O/E	1.03	(0.26)

If agreement for voicing were a necessary condition or consequence of retroflex consonant harmony then we would expect pairs of plosives that agree in retroflexion to also agree in voicing. The data in Table 13 reveal that pairs of plosives that agree in both retroflexion and voicing are in fact favoured in every case. O/E ratios range from 1.29 for Kui to 3.05 for Kurux. Significantly, however, pairs that agree in retroflexion but disagree in voicing are not disfavoured; they occur more-or-less as expected in every case. O/E ratios range from 0.77 for Kurux to 1.27 for Parji. Thus, there is nothing to suggest that voicing plays a role in retroflex consonant harmony in these languages. This conclusion is further corroborated by the historical-comparative data reviewed in §3.1.4 above. Cases of retroflex consonant harmony without agreement for laryngeal features are not hard to find. Examples include Malto /ʈuɖ/ ‘to smear’ (cf. Telugu /tuɖutʃu/), Konda /ɖaʈam/ ‘many, thick’ (cf. Telugu /daʈʈamu/) and Parji /ʈõːɖ-/ ‘to draw water’ (cf. Telugu /tõːɖu/).

In sum, retroflex consonant harmony is characteristic of most North and South-Central Dravidian languages and of Central Dravidian Parji. Retroflex consonant harmony in Dravidian exhibits the following typological properties: (i) it is root-internal; (ii) it is regressive (or possibly bidirectional); (iii) it is sensitive to similarity in terms of manner of articulation so that it holds between two plosives but not between plosives and sonorants; (iv) it is not blocked by intervening segments; and (v) it does not require or entail agreement for laryngeal features.

3.2 Indo-Aryan

The pattern of retroflex consonant harmony evident in Dravidian also applies to the majority of New Indo-Aryan languages. This section begins by reviewing evidence that retroflex consonant harmony did not occur in Old-Indo-Aryan Sanskrit (§3.2.1) or, for the most part, in Middle Indo-Aryan languages (§3.2.2), though the seeds of it are evident in Prakrit. After reviewing evidence from Old and Middle Indo-Aryan, the study turns to the New Indo-Aryan languages and demonstrates that retroflex consonant harmony has applied to contemporary Panjabi (§3.2.3) and to most other New Indo-Aryan languages of the Northwestern, Northern, Central and Eastern zones but not to those of the Southern zone or Sinhalese-Maldivian group (§3.2.4).

3.2.1 Sanskrit (Old Indo-Aryan)

Sanskrit is representative of the Old-Indo-Aryan period (c. 1500 BC – 600 BC) and is well known for its pattern of retroflex assimilation known as "nati" or "n-retroflexion". This section reviews evidence that Sanskrit n-retroflexion is not a true case of consonant harmony, as defined in the present study (§3.2.1.1), and that retroflex consonant harmony of the kind found in Dravidian did not occur in Sanskrit (§3.2.1.2).

The consonant phonemes of Sanskrit are listed in (24).

(24) Consonant phonemes of Sanskrit (Whitney, 1993 [1889]; Cardona, 2003)

LAB	DEN	RET	PAL	VEL	GLOT
p	t	ʈ	ʧ	k	
p^h	t^h	$ʈ^h$	$ʧ^h$	k^h	
b	d	ɖ	ʤ	g	
b^h	d^h	$ɖ^h$	$ʤ^h$	g^h	
	s	ʂ	ʃ		(h) ɦ
m	n	ɳ	ɲ	ŋ	
	l				
	r				
v			j		

As shown in (24), Sanskrit had a rich system of coronal consonants that included stops, fricatives and nasals at three places of articulation: dental, retroflex and palatal. The rhotic /r/ is described variously as dental, alveolar or retroflex (Cardona, 2003, p. 109). Both /r/ and /l/ have syllabic counterparts (/r̩, l̩/) that are traditionally treated as part of the vowel system. The glottal continuant, commonly transliterated *h*, was in fact breathy voiced /ɦ/. The segment commonly transliterated *ḥ* corresponds to orthographic *visarga*, which represents a voiceless glottal fricative (IPA [h]) that can be treated as an allophone of /s/.

3.2.1.1 Sanskrit n-retroflexion

Sanskrit n-retroflexion is a case of progressive assimilation triggered by retroflex continuants and targeting dental nasals: dental /n/ is realized as retroflex /ɳ/ when it follows /ʂ/ or /r/ (or its vocalic counterpart /r̩/, whether short or long). The trigger and target may be adjacent (25)(a) or non-adjacent. Non-adjacent assimilation can extend across an intervening vowel (25)(b), consonant (25)(c) or longer string of segments (25)(d). It also extends across morpheme boundaries separating roots from derivational or inflectional affixes and even across word boundaries in some cases (Whitney, 1993 [1889], pp. 65–66). The examples in (25) and all subsequent examples cited in this section are adapted from Schein & Steriade (1986) and Hansson (2001; 2010).

(25) Sanskrit n-retroflexion

a.	iʂ-naː-	→ iʂ-ɳaː-	'seek (present stem)'
	puːr-na-	→ puːr-ɳa-	'filled (passive participle)'
b.	ʧakʂ-aːna-	→ ʧakʂ-aːɳa-	'see (middle participle)'
	pur-aːna-	→ pur-aːɳa-	'fill (middle participle)'
c.	vr̩k-na-	→ vr̩k-ɳa-	'cut up (passive participle)'
	gr̩b^{h}-naː-ti	→ gr̩b^{h}-ɳaː-ti	'seizes (3Sg active)'
d.	kʂub^{h}-aːna-	→ kʂub^{h}-aːɳa-	'quake (middle participle)'
	kr̩p-a-maːna-	→ kr̩p-a-maːɳa-	'lament (middle participle)'

The pattern of assimilation is purely progressive, never regressive, and it targets only the first dental nasal to the right of the trigger. In other words it is non-iterative; the derived retroflex nasal in the sequence ʂ/r…ɳ…n does not induce retroflexion of any subsequent dental nasals, nor does it allow the preceding ʂ/r to do so, as shown in (26).

(26) N-retroflexion targets only the first nasal to the right of the trigger

pra-ɳinaːja	(*pra-ɳi-ɳaːja)	'lead forth' (/niː-/ 'lead')
kr̩ɳ-vaːna	(*kr̩ɳ-vaːɳa)	'make (middle participle)'

Assimilation is also blocked when any other coronal consonant intervenes between the trigger and target, whether it is dental, retroflex or palatal, obstruent or sonorant. The only exception is the palatal glide /j/, which behaves more like a vowel in this case.[23]

(27) N-retroflexion is blocked by intervening coronals

mr̩d-naː-	(*mr̩d-ɳaː-)	'be gracious (present stem)'
marʤ-aːna-	(*marʤ-aːɳa-)	'wipe (middle participle)'
kr̩t-a-maːna-	(*kr̩t-a-maːɳa-)	'cut (middle participle)'
kʂved-aːna-	(*kʂved-aːɳa-)	'hum (middle participle)'

Sanskrit n-retroflexion displays several other curious properties. For example, assimilation only occurs if /n/ is immediately followed by a non-liquid sonorant (i.e., a vowel, glide or nasal) but not if it is in final position (28)(a) or if it is immediately followed by an obstruent or liquid (28)(b).

(28) N-retroflexion applies only if /n/ is immediately followed by a non-liquid sonorant

a.	braɦman	(*braɦmaɳ)	'brahman (VocSg)'	cf. braɦmaɳ-i	(LocSg)
b.	tr̩-n-t-te	(*tr̩-ɳ-t-te)	'split (3Pl middle)'	cf. tr̩-ɳa-t-ti	(3Sg active)

The conditions illustrated in (28) might not be true properties of n-retroflexion *per se*, but might derive from the interaction of n-retroflexion with other aspects of Sanskrit phonology, as argued by Schein & Steriade (1986). For example, the non-application of n-

[23] Strictly speaking, it is impossible to determine whether /ʂ/ and /r/ serve as blockers since they also serve as triggers. In a sequence such as ʂ/r…ʂ/r…ɳ, the retroflexion of the nasal would be conditioned by the medial ʂ/r and it is impossible to say whether the initial ʂ/r also has any effect or whether its effect is blocked.

retroflexion to final nasals (as in (28)(a)) might stem from the neutralization of nasal place in that environment; final nasals are invariably labial /m/ in inflectional affixes and dental /n/ elsewhere (e.g., a-gam-∅ → agan 'go (2Sg aorist)', but a-gam-a-m → agamam 'go (1Sg aorist)'). The non-application of n-retroflexion to nasals immediately followed by stops (as in (28)(b)) might stem from the independent condition requiring regressive place assimilation in nasal-stop sequences. This condition over-rides n-retroflexion so that we find /tr̩-n-t-te/ 'split (3Pl middle)' with homorganic dental /-n-t-/ instead of */-ɳ-t-/, which would be the expected output of n-retroflexion.[24] Before /l/, /n/ is realized as a nasalized lateral and before non-syllabic continuants (/s, ʂ, ʃ, r/ and /ɦ/) it is realized as nasalization of the preceding vowel (i.e., orthographic *anuswāra*). All of these conditions over-ride and bleed n-retroflexion, thereby restricting it to cases where /n/ is immediately followed by a non-liquid sonorant.

Sanskrit n-retroflexion has long been cited as a classic example of coronal consonant harmony. In spite of this, Hansson notes that it "stands out like a sore thumb" with respect to the cross-linguistic typology of consonant harmony systems (2001, pp. 81, 238; cf. 2010, pp. 179–193). N-retroflexion violates the most significant typological properties exhibited time and again by other consonant harmony systems. For example, most consonant harmonies exhibit similarity effects; the class of triggers and targets are those that are most similar to each other in terms of phonological features or phonetic properties. In the case of n-retroflexion the class

---

[24] In principle, n-retroflexion and the homorganic restriction on nasal-stop clusters could both be satisfied through *progressive* assimilation within the nasal-stop sequence (i.e., ʂ/r…nt → ʂ/r…ɳʈ). Progressive retroflex assimilation does occur in nasal-stop sequences but only in cases where the retroflex nasal is non-derived (e.g., pʰaɳ-ta- → pʰaɳ-ʈa- 'spring (passive participle)'). Progressive assimilation of this kind is never triggered by derived retroflex nasals such as those produced through n-retroflexion. As a result, assimilation in nasal-stop sequences is predominantly regressive.

of triggers and targets are distinct (/ʂ, r/ vs. /n/). If anything, we might expect /s/ to serve as a target, given that /ʂ/ is a trigger, or we might expect /ɳ/ to serve as a trigger, given that /n/ is a target. This is not the case. Sanskrit n-retroflexion clearly passes over the most similar of potential candidates in favour of other less similar segments. According to Hansson, this is entirely unheard of in other consonant harmony systems.

Another important typological anomaly of Sanskrit n-retroflexion is the blocking effect exhibited by coronal consonants that intervene between the trigger and target. While blocking (or 'opacity') effects are quite common in vowel harmony and vowel-consonant harmony systems, Hansson finds that they are exceedingly rare in consonant harmony systems. In consonant harmony, those segments that do not participate as triggers or targets are typically transparent or neutral with respect to harmony. This finding was also affirmed in an independent study by Rose & Walker (2004).

Other typological anomalies of n-retroflexion identified by Hansson include the strictly progressive direction and the domain, which potentially extends across word boundaries at a phrasal level. Cross-linguistically, consonant harmony systems typically exhibit regressive assimilation, not progressive. Hansson argues that the rare cases of progressive assimilation can be attributed to stem control. As such, they are not true cases of progressive assimilation but rather "inside-out" assimilation, with segments in affixes (whether prefix or suffix) assimilating to those in roots. When it comes to the domain of harmony, consonant harmony can apply within morphemes or across morpheme boundaries. However, there are no other known cases in which harmony holds across word boundaries, as reported for n-retroflexion.

The most important typological anomalies of Sanskrit n-retroflexion identified by Hansson (2001, pp. 240–241; 2010, p. 191) are summarized in (29).

(29) Sanskrit n-retroflexion as consonant harmony—typological anomalies (Hansson, 2010)

a. Segmental opacity: For a particular class of segments to block the propagation of harmony is virtually unattested in consonant harmony systems.

b. Triggers vs. targets: In no other consonant harmony system is the set of triggers disjoint with that of targets; consonant harmony always respects relative similarity.

c. Directionality: Perseveratory [= progressive] directionality which does not emerge from constituent structure (or other faithfulness effects) is otherwise unattested (or at best extremely rare) in consonant harmony systems.

d. Harmony domain: In no other consonant harmony system does the assimilation apply at a phrasal (or clitic-group) level, reaching across word boundaries.

In light of its unusual properties, Hansson argues that Sanskrit n-retroflexion is not a true case of consonant harmony; it is not the product of the same assimilatory mechanisms that underlie (true) consonant harmony systems. Rather, it is more akin to vowel or vowel-consonant harmony with which it shares more typological properties, and which arguably involve different mechanisms of assimilation (Hansson, 2001; 2010; Rose & Walker, 2004).

3.2.1.2 Synchronic co-occurrence patterns in Sanskrit

Assuming that Sanskrit n-retroflexion is not a true case of consonant harmony, we may ask: did retroflex consonant harmony of the type attested in Malto and other Dravidian languages also occur in Sanskrit? To this the answer is clearly “no”. The co-occurrence pattern of dental and retroflex plosives in Sanskrit was essentially the same as that of Proto-Dravidian in (2). In the earliest Sanskrit records, retroflex consonants did not occur word-initially. The few forms with word-initial retroflex consonants listed in most Sanskrit dictionaries are mostly later innovations that appear in texts only from the fifth and sixth centuries onwards (Masica, 1991, p. 157; Jain, 1934, pp. 57–58). Table 14 shows observed counts and O/E ratios for word-initial $C_1V(N)C_2$ sequences in Apte's (1957-1959) Sanskrit dictionary, where C_1 and C_2 are coronal plosives or retroflex sonorants (i.e., /ɳ/) and N is a homorganic nasal.

Table 14 Sanskrit coronal plosives and retroflex sonorants in #$C_1V(N)C_2$ sequences (Apte, 1957–1959; n = 393)

C_1\ C_2		t, tʰ, d, dʰ	ʈ, ʈʰ, ɖ, ɖʰ	ɳ
t, tʰ, d, dʰ	O	257	101	19
	O/E	1.03	0.92	1.04
ʈ, ʈʰ, ɖ, ɖʰ	O	2	14	0
	O/E	0.19	(2.99)	(0.00)

The Sanskrit data in Table 14 exhibits the same general pattern found in Dravidian languages that lack retroflex consonant harmony (cf. Telugu, Gadaba and Tamil in Table 12). Harmonic Ṭ-Ṭ sequences are over-attested (O/E = 2.99) owing to low observed counts for word-initial retroflexes (only 16 out of 393 headwords in the data set), which in turn lead to extremely low expected frequencies for Ṭ-Ṭ sequences (E = 4.7). Significantly, however, disharmonic T-Ṭ sequences are well attested (O = 101, O/E = 0.92) and this is a more reliable indicator that harmony was not enforced (cf. discussion in §1.4).

A similar pattern is found for Sanskrit coronal fricatives as shown in Table 15.

Table 15 Sanskrit fricatives in $\#C_1V(N)C_2$ sequences (Apte, 1957–1959; n = 322)

C_1\ C_2		s	ʂ	ʃ
s	O	124	27	61
	O/E	1.18	0.55	1.04
ʂ	O	0	7	0
	O/E	(0.00)	(4.35)	(0.00)
ʃ	O	35	40	28
	O/E	0.69	1.69	0.98

Once again, harmonic Ṣ-Ṣ sequences are over-attested (O/E = 4.35) owing to an extremely low expected frequency (E = 1.6). It might be tempting to see the seeds of retroflex consonant harmony in the fact that word-initial /ʂ/ occurs only before another /ʂ/ in C_2 position (in words where both C_1 and C_2 are coronal fricatives) and that disharmonic S-Ṣ is somewhat under-attested (O/E = 0.55). However, this is doubtful considering that observed counts of Ṣ-Ṣ are extremely low (7 out of 322 words in the data set) and that all instances appear to be derivatives of the same root (e.g., /ʂaʂ/ 'six', /ʂaʂʈiḥ/ 'sixty', /ʂaʂʈ^{h}a/ 'sixth', /ʂaʂʈika/ 'bought with sixty', etc.). Although they are under-attested, S-Ṣ sequences do not exhibit the categorical or near-categorical absence that is characteristic of disharmonic sequences in other cases of consonant harmony. Moreover, disharmonic Š-Ṣ sequences are actually over-attested (O/E = 1.69).

In sum, Old Indo-Aryan Sanskrit exhibited a form of retroflex assimilation known as n-retroflexion. However, this pattern of assimilation exhibits a number of typological anomalies that set it apart. The weight of evidence suggests that it is not a product of the same assimilatory mechanisms that underlie consonant harmony systems. Evidence reviewed here indicates that retroflex consonant harmony of the kind found in Malto and other Dravidian languages did not occur in Old Indo-Aryan Sanskrit, either between plosives or sibilants.

3.2.2 Pāli & Prakrit (Middle Indo-Aryan)

For the most part, retroflex consonant harmony did not apply to languages of the Middle Indo-Aryan period (MIA, c. 600 BC – 1000 AD), which include Pāli and Prakrit, although the early stages of harmony can be seen in Prakrit. All MIA languages and dialects preserved the OIA contrast between dental, retroflex and palatal stops but most neutralized coronal place contrasts between fricatives. Thus, both Pāli and Prakrit neutralized the three-way coronal fricative system of OIA to a single dental series (OIA /s, ʂ, ʃ/ > MIA /s/). In terms of sonorants, Pāli preserved the distinction between dental and retroflex nasals (/n/ vs. /ɳ/) and also developed retroflex laterals /ɭ/ and /ɭʰ/ (< OIA /ɖ/ and /ɖʰ/, intervocalically). Prakrit, on the other hand, merged all coronal nasals to retroflex /ɳ/.[25]

Like OIA Sanskrit, Pāli shows no evidence of retroflex consonant harmony between plosives. For the most part Prakrit also lacks retroflex consonant harmony although the seeds of it are evident. Table 16 and Table 17 show observed and O/E values for word-initial $C_1V(N)C_2$

---

[25] It is not clear whether this merger was real or artificial, i.e., whether it occurred in speech or whether it was partly the product of orthographic conventions. There is evidence to suggest that the merger to retroflex [ɳ] was real enough in non-initial positions, but it is possible that word-initial coronal nasals were actually dental [n] (Schwarzschild, 1973; Masica, 1991, p. 182). At any rate it is clear that the contrast between /n/ and /ɳ/ was neutralized. Only /ɳ/ occurs in Turner’s wordlist (1969, pp. 25–56), whether initial or non-initial, and this is the source of data for the present study.

sequences in Pāli and Prakrit (respectively), where C_1 and C_2 are coronal plosives or retroflex sonorants and N is a homorganic nasal.

Table 16 Pāli coronal plosives and retroflex sonorants in #C_1V(N)C_2 sequences (Pali Text Society, 1921–1925; n = 202)

C_1\ C_2		t, tʰ, d, dʰ	ʈ, ʈʰ, ɖ, ɖʰ	ɳ, ɭ, ɭʰ
t, tʰ, d, dʰ	O	123	37	34
	O/E	0.99	0.99	1.04
ʈ, ʈʰ, ɖ, ɖʰ	O	6	2	0
	O/E	1.17	(1.29)	(0.00)

Table 17 Prakrit coronal plosives and retroflex sonorants in #C_1V(N)C_2 sequences (Turner, 1969; n = 337)[26]

C_1\ C_2		t, tʰ, d, dʰ	ʈ, ʈʰ, ɖ, ɖʰ	ɳ, ɳʰ
t, tʰ, d, dʰ	O	84	70	69
	O/E	0.99	0.90	1.15
ʈ, ʈʰ, ɖ, ɖʰ	O	0	21	7
	O/E	0.00	2.14	0.93
ɳ, ɳʰ	O	44	27	15
	O/E	1.35	0.90	0.65

[26] Prakrit is a cover term for a collection of Middle Indo-Aryan vernacular dialects. In Turner's data, these include Śaurasenī, Paiśācī, Ardhamāgadhī, Māgadhī, Jaina Māgadhī, Mahārāṣṭrī and Jaina Mahārāṣṭrī (1969, p. vii).

Harmonic Ṭ-Ṭ sequences occur more frequently than expected in both languages. However, in each case disharmonic T-Ṭ sequences are also well attested. If consonant harmony applied to these languages we would expect to find disharmonic T-Ṭ sequences largely avoided. Thus, on the whole, neither language exhibits a clear pattern of retroflex consonant harmony.

Although Prakrit does not exhibit a clear pattern of retroflex consonant harmony, the early stages of harmony are evident. This can be seen in the larger number of observed Ṭ-Ṭ sequences (O = 21), some of which are the product of dialectal variation between harmonic and disharmonic forms of selected roots. Both harmonic and disharmonic variants are attested for many Prakrit roots. Some examples are listed in (30). Reference numbers in the right-most column of (30) refer to etymological groups in Turner's (1962–1966) *Comparative Dictionary of Indo-Aryan Languages* (CDIAL).

(30) Variable retroflex consonant harmony in selected Prakrit roots (Turner, 1962–1966)

	Prakrit		OIA	CDIAL
'screen; hedge'	ʈaʈʈiː-	taʈʈiː-	*traʈʈa-	5990
'is broken'	ʈuʈʈaï	tuʈʈaï	truʈjati	6065
'burnt'	ɖaɖɖ^{h}a-	daɖɖ^{h}a-	dagd^{h}a	6121
'stick'	ɖãɖa-	dãɖa-	daɳɖa-	6128
'bitten, stung'	ɖaʈʈ^{h}a-	daʈʈ^{h}a-	daʂʈa	6243
'cold; fixed, firm'	ʈ^{h}aɖɖ^{h}a-	t^{h}aɖɖ^{h}a-	stabd^{h}a	13676

The examples in (30) suggest that retroflex consonant harmony may have operated to a limited extent in some Prakrit dialects, though it was clearly not a widespread property of MIA as evidenced by the co-occurrence patterns in Table 16 and Table 17.

3.2.3 Panjabi (New Indo-Aryan)

The beginnings of retroflex consonant harmony evident in MIA Prakrit are carried to their full extent in many NIA languages. Panjabi, a NIA language spoken primarily in the state of Panjab in northwest India, is representative of this trend. The consonants of Panjabi are shown in (31).

(31) Consonant phonemes of Panjabi (Bhatia, 1993; Malik, 1995; Shackle, 2003)

LAB	DEN	ALV	RET	PAL	VEL	GLOT
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)	
(f)		s		ʃ	(x)	h
		(z)			(ɣ)	
m		n	ɳ			
		l	(ɭ)			
		r	ɽ			
w				j		

Panjabi has lost the voiced aspirated stops of OIA and developed a tonal system in their place. The voiced aspirates are still preserved in orthography but are now pronounced as their unaspirated counterparts with accompanying pitch contours on neighbouring vowels. In much of the literature on Panjabi, phonemic transcriptions mirror the orthography by using voiced aspirated characters to represent what is essentially a tonal contrast. Since this convention is followed in most of the data sources employed for the current study I will continue to use it here. For this reason the voiced aspirates are included in (31) but enclosed in parentheses to mark their special status. All other segments enclosed in parentheses are marginal to the system and not distinguished by all speakers or dialects.

While Panjabi has lost contrastive retroflexion in the fricative class it has maintained it among plosives and nasals and extended it to liquids. Most of the intervocalic (singleton) stops of OIA were lost altogether through a process of lenition. However, intervocalic retroflex plosives were preserved as the retroflex flap /ɽ/, which subsequently gained independent phonemic status (Jain, 1934, pp. 58–59, §138). The retroflex lateral /ɭ/ is not phonemic in all dialects and is not distinguished in the orthography or in any of the data sources employed for the current study. Thus, it will not be discussed further.[27]

The retroflex sonorants /ɳ/ and /ɽ/ are subject to the same phonotactic restriction that applied to most retroflexes in OIA: they do not occur word-initially. However, retroflex plosives, once avoided word-initially, now appear frequently in that position. Commenting on the historical development of Panjabi retroflex plosives from OIA dentals, particularly in word-initial position, Jain points out a "tendency to cerebralise [= retroflex] a dental stop occurring in the vicinity of another cerebral [= retroflex] stop", where "in the vicinity" means nearby but non-adjacent (1934, pp. 89, §171). In the following sub-section I present the results of a study examining synchronic co-occurrence restrictions on dental and retroflex plosives in Panjabi. The results of the study suggest that the pattern of consonant harmony observed by Jain (1934) is quite extensive in the language and exhibits the same typological properties that characterize retroflex consonant harmony in Malto and other Dravidian languages.

---

[27] Many details concerning retroflex /ɭ/ are unclear in the literature. For example, Bhatia (1993) and Shackle (2003) both report that /ɭ/ is phonemic only in the standard Majhi dialect, but Malik (1995), explicitly restricting his study to that dialect, has [ɭ] as an allophone of dental /l/. The situation is not helped by the fact that retroflex /ɭ/ is not distinguished in the orthography. Moreover, it does not occur in the dialect described by Jain (1934), which is my principal source for Panjabi historical phonology.

3.2.3.1 Synchronic co-occurrence patterns in Panjabi

In order to investigate the extent of retroflex consonant harmony in Panjabi, a study was conducted based on data from Goswami's (2000) Panjabi-English dictionary. Once again, the study counted headwords containing word-initial #C_1V(N)C_2 sequences in which C_1 and C_2 are coronal plosives or retroflex sonorants and N is a homorganic nasal. The results are shown in Table 18. Goswami's dictionary does not distinguish retroflex /ɭ/. Thus, the class of retroflex sonorants is limited to /ɳ/ and /ɽ/ and a few instances of /ɽʰ/ in this case.

Table 18 Panjabi coronal plosives and retroflex sonorants in #C_1V(N)C_2 sequences (Goswami, 2000; n = 233)

C_1\ C_2		t, tʰ, d, dʰ	ʈ, ʈʰ, ɖ, ɖʰ	ɳ, ɽ, ɽʰ
t, tʰ, d, dʰ	O	70	5	70
	O/E	1.61	0.10	1.41
ʈ, ʈʰ, ɖ, ɖʰ	O	0	78	10
	O/E	0.00	2.49	0.33

The results in Table 18 suggest a robust co-occurrence restriction on dental and retroflex plosives in Panjabi. Sequences that agree in retroflexion or non-retroflexion are overwhelmingly preferred, whether Ṭ-Ṭ (O/E = 2.49) or T-T (O/E = 1.61). More importantly, sequences that disagree in retroflexion are either categorically absent, as in the case of Ṭ-T (O/E = 0.00), or nearly so, as in the case of T-Ṭ (O/E = 0.10). Of the five exceptional disharmonic T-Ṭ sequences identified in Goswami (2000), three are listed as having harmonic variants, as shown in (32)(a). The two remaining exceptions are both derivatives of a single root and are most likely *tatsamas* (i.e., unassimilated Sanskrit loans), as shown in (32)(b).[28]

(32) Exceptions to retroflex consonant harmony in Goswami's (2000) Panjabi dictionary

a. Variation between disharmonic and harmonic forms of the same root (T-Ṭ ~ Ṭ-Ṭ)

toʈʈaː	'loss, deficiency'	cf.	ʈoʈaː	'deficiency, loss'
toɖaː	'camel's young'	cf.	ʈoɖaː	'young camel'
danɖ	'punishment'	cf.	ɖanɖ	'punishment'

b. Unassimilated Sanskrit loanwords (*tatsamas*)

taʈʈ	'coast, bank, shore'	cf. Skt.	taʈa-	'slope, shore'
taʈast^{h}	'neutral' (lit. 'standing on a bank or shore')	cf. Skt.	taʈa-	'slope, shore'

The results in Table 18 indicate that Panjabi is clearly affected by a tendency toward retroflex consonant harmony between coronal plosives. They also indicate that harmony is limited to the class of plosives. Retroflex sonorants do not participate. Disharmonic T-Ṛ sequences are actually preferred (O/E = 1.41) over harmonic Ṭ-Ṛ sequences (O/E = 0.33). Thus, Panjabi exhibits the same similarity effect observed in Malto and the other Dravidian languages

---

[28] Within the Indo-Aryan tradition, a distinction is commonly made between *tadbhavas* and *tatsamas*. The term *tadbhava* is Sanskrit for 'originating from that'. It refers to NIA words that have been inherited from Sanskrit through the normal channels of transmission and, therefore, subjected to systematic sound changes. The term *tatsama* is Sanskrit for 'the same as that'. It refers to Sanskrit loanwords borrowed directly into the NIA languages with little or no phonological adaptation (at least orthographically, if not in pronunciation). With respect to the Panjabi words in (32)(b), McGregor (1993, p. 434) clearly identifies the corresponding words in Hindi as *tatsamas*.

reviewed in §3.1: retroflex consonant harmony holds between coronal plosives but not between plosives and sonorants even though retroflexion is contrastive in both manner classes.

3.2.3.2 Historical-comparative evidence of consonant harmony

Historical and comparative data confirm that retroflex consonant harmony was limited to plosives in Panjabi; dental plosives were always the targets of assimilation and retroflex plosives always the triggers. They also reveal a regressive direction of assimilation and general absence of blocking effects. These effects can be seen quite clearly by comparing the Panjabi examples in (33) with cognates from MIA Prakrit and OIA Sanskrit.[29]

(33) Evidence of retroflex consonant harmony in Panjabi

		Panjabi	MIA	OIA	CDIAL
a.	'screen'	ʈaʈʈiː	ʈaʈʈiː-, taʈʈiː-	*traʈʈa-	5990
	'to break'	ʈuʈʈɳaː	ʈuʈʈaï, tuʈʈaï	truʈjati	6065
	'grasshopper'	ʈiɖɖaː	tiɖɖa-	*tiɖɖa-, *triɖɖa-	6024
	'crooked'	ʈeɖʰaː	—	*treɖɖʰa-	6071
	'plug'	ɖaʈʈaː	-daʈʈa-	*draʈʈ-	6618

[29] Unless otherwise noted, the Panjabi data here and elsewhere are taken from Jain (1934) and Goswami (2000) and the Prakrit and Sanskrit data are primarily from Turner (1962–1966). OIA forms prefixed with an asterisk (*) are hypothetical forms reconstructed by Turner on the basis of comparative data. All other forms are attested in Sanskrit. The OIA and MIA forms may or may not be the precise ones from which the corresponding Panjabi words have developed in every case. The point is only to show that there are historical cognates that have initial dental stops where Panjabi has retroflexes.

b.	‘to open’	ʈaɖɖnaː	taɖɖaï	tardati	5721
	‘lamp stand’	ɖiũːʈ	—	diːpavarti	6354
	‘frog’	ɖaɖɖuː	ɖeɖɖura, daddura	dardura-	6198
	‘strong’	ɖaːɖɖ^{h}aː	daɖ^{h}a-	daːrɖ^{h}ja-, dr̩ɖ^{h}a-	6508
c.	‘pony’	ʈaʈʈuː	*ʈaʈʈu-, *taʈʈu-	—	5440
	‘tap, spout’	ʈuːʈiː	tũɖa	tuɳɖa	5853
	‘stick’	ɖanɖaː	ɖãɖa-, dãɖa	daɳɖa-	6128
	‘vessel for curd’	ɖahinɖi	—	dad^{h}ib^{h}aːɳɖa	6149
	‘to be kind’	ʈuʈʈ^{h}naː	tuʈʈ^{h}a-	tuʂʈa-	5895
	‘fallen’	ɖ^{h}aːʈ^h	*d^{h}aʈʈ^{h}a-, d^{h}att^{h}a-	*d^{h}vaʂʈa- < d^{h}vasta-	6896

In each of the examples in (33)(a) C_1 corresponds to an OIA dental + r sequence (either attested or reconstructed), while in (33)(b) C_2 corresponds to an OIA r + dental or r + retroflex sequence. Since rhotics are one of the most common sources of retroflexion it is reasonable to question whether these are true cases of consonant harmony holding between stops or whether retroflexion was simply spread from the rhotic to surrounding segments. There is evidence that the rhotic is not responsible for the harmony, although is some cases it might be responsible for the retroflexion of the consonant in C_2 position. First of all, the examples in (33)(c) do not stem from forms with rhotics. The only thing these examples share with those of (33)(a) and (b) is the presence of a retroflex stop non-initially (in the MIA cognate if not in the OIA cognate). Thus, if all of the examples in (33) are the result of a single pattern of sound change (the null-hypothesis) then the word-initial retroflex stops must be the product of harmony with the non-initial stops (i.e., T-Ṭ → Ṭ-Ṭ).

Secondly, with respect to the historical influence of rhotics on dentals, Jain reports: "Panjābī is a non cerebralising [= non-retroflexing] dialect" (1934, p. 87). That is, OIA dentals occurring in the vicinity of rhotics generally remained dental in Panjabi and were not subject to retroflexion. Western Panjabi typically retains OIA rhotics in clusters where Eastern Panjabi has lost them. Either way, both varieties have retained OIA dentals when they occurred before /r/, as shown in (34).

(34) OIA dentals preserved before /r/ in Panjabi (cf. Jain, 1934, p. 81)

		Eastern Panjabi	Western Panjabi	OIA
a.	'to drive away'	taːhnaː	traːhɳaː	traːsajati
	'to break'	toːɽnaː	troɽnaː	troːʈajati
	'three'	tinn	trai	triːɳi, trajaḥ
	'price'	damm	—	dramma-
	'grape'	daːk^h	—	draːkṣaː
b.	'sub caste'	goːt	goːttar	goːtra-
	'leopard'	ʧittaː	ʧitraː	ʧitraka-
	'daughter's son'	dohtaː	doːhtraː	dauɦitra-
	'son'	putt	puttar	putra-
	'thread'	sut	suːttar	sutra-
	'sickle'	daːttiː	daːtriː	daːtra-
	'ringworm'	dadd	daddar	dadru-
	'sleep'	niːd	niːndar	nidraː

c.	'to cross'	tarnaː	—	tarati
	'slanting'	tirʧʰaː	—	tiraʃʧa
	'to walk'	turnaː	—	turati
	'distant'	duːr	—	duːra
	'to place'	dʰarnaː	—	dʰarati

The examples in (34) demonstrate that dental stops remain dental before /r/ regardless of whether the dental + r sequence occurred word-initially (34)(a), or non-initially (34)(b), or whether the rhotic followed at a distance (34)(c). We can only conclude, along with Jain (1934, p. 81), that "r after a dental stop does not cause cerebralisation [= retroflexion]" and that in examples like those in (33)(a) (e.g., ʈuʈʈnaː 'to break' < Pk. tuʈʈa- < Skt. truʈjati) the retroflexion of the initial stop is "due to the following -ʈʈ-" and not to the presence of /r/.

In most cases, dentals were also retained in Panjabi when they were preceded by a rhotic, as shown in (35).

(35) OIA dentals preserved after /r, r̩/ in Panjabi (cf. Jain, 1934, pp. 87–88)

		Panjabi	OIA
a.	‘company’	saːt^{h}	saːrt^{h}a-
	‘fourth’	ʧautt^{h}aː	ʧaturt^{h}a-
	‘ass’	gad^{h}aː	gardab^{h}a-
	‘to jump’	kuddanaː	kuːrdati
	‘cloud’	baddal	vaːrdala-
	‘to break wind’	paddanaː	pardati
	‘ginger’	aːddaː	aːrdra-
	‘half’	add^{h}	ard^{h}a-
b.	‘constellation’	k^{h}ittiː	kr̩ttikaː
	‘done’	kiːttaː	kr̩ta
	‘clarified butter’	g^{h}eoː	g^{h}r̩ta-
	‘dead’	moːeaː	mr̩ta
	‘heart’	hiːaː	hr̩daja-
c.	‘brother’	b^{h}aːiː	b^{h}raːtr̩-
	‘killed’ (?)	maːreaː	maːrita
	‘weeping’	roːn	roːdana-
	‘first’	paihllaː	prat^{h}illa- (cf. prat^{h}ama-)

The examples in (35)(a) demonstrate that dental stops typically remain dental immediately after consonantal /r/, while those in (35)(b) demonstrate the same point with respect to syllabic /r̩/. Lastly, the examples in (35)(c) demonstrate that dentals also remained dental when preceded by /r/ at a distance. Notice that in (35)(b) and (35)(c) the dental stop is often lost altogether in Panjabi. This is evidence that it was not subject to retroflexion. Recall that most OIA single intervocalic stops were lost through lenition in Panjabi. However, single intervocalic retroflex stops were preserved as retroflex flaps. Thus, as Jain has pointed out: "Here the dentals remained dental which subsequently disappeared. If they had become cerebral [= retroflex] they should have appeared as [ɽ, ɽʰ]" (Jain, 1934, p. 87).

There are, however, some cases where OIA rhotics did induce retroflexion on an immediately following dental stop. Examples are provided in (36).

(36) Panjabi retroflex stops from OIA dentals after /r, r̩/ (cf. Jain, 1934, p. 88)

	Panjabi	**OIA**
'cowrie shell'	kauɖɖiː	kapardikaː
'to leave'	ʧʰaɖɖᵃnaː	ʧardati
'frog'	ɖaɖɖuː	dardura-
'beginning'	muɖɖʰ	muːrdʰan
'earth'	miʈʈiː	mr̩ttikaː

The examples in (36) demonstrate that OIA r/r̩ + dental stop sequences did produce retroflex stops in some cases. Jain points out that examples of this kind tend to show up with retroflexion in most other NIA languages as well. This suggests that they represent a very early dialectal development within Indo-Aryan and not something unique to Panjabi. The variable treatment of dentals after /r/ probably goes back to dialectal variation in MIA. Commenting on phonological correspondences between OIA and MIA, Masica notes that the OIA sequence r + dental "sometimes yields a retroflex, sometimes a dental geminate, rather unpredictably"

(1991, p. 176). This variation shows up in Panjabi (and elsewhere) in the form of doublets (i.e., minimal or near-minimal pairs) differing primarily in their treatment of OIA r + dental sequences. Some examples from Panjabi are listed in (37).

(37) Doublets resulting from variable treatment of OIA dentals after /r, r̩/ (Jain, 1934, p. 88)

		Panjabi	OIA
a.	'to spin'	kattanaː	kr̩natti
	'to cut'	kaʈʈanaː	kartati
b.	'thing, matter'	baːt	vaːrttaː
	'road'	baːʈ	vartma-, vartiḥ
c.	'to increase'	badʰnaː	vardʰate
	'to cut'	baɖɖʰanaː	vardʰati

Jain suggests that the variable treatment of r + dental sequences in examples like those in (37) may have been motivated by the need to preserve a semantic contrast (1934, p. 88). Whatever the case may be, it is clear that within the particular stream of development that ultimately produced Panjabi, rhotics did not induce retroflexion on preceding dentals (dental + r) but did induce retroflexion of following dentals (r + dental) in some cases. Thus, in examples such as those in (33)(b) (e.g., ʈaɖɖɳaː 'to open' < tardati), the presence of a rhotic in OIA may have induced retroflexion on the dental in C_2 position but cannot be responsible for the retroflexion of the dental in C_1 position. The retroflexion of the word-initial dental must be the result of consonant harmony induced by the plosive in C_2 position (e.g., t…rd → t…ɖɖ → ʈ…ɖɖ) as sketched in (38).

(38) Development of retroflex consonant harmony from OIA C_1…r C_2 sequences

	Panjabi				OIA
‘to open’	ʈaɖɖ-ɳaː	<	taɖɖ-	<	tardati
‘frog’	ɖaɖɖuː	<	*daɖɖ-	<	dardura-

Granted that the rhotics /r, r̩/ did not trigger retroflex harmony in Panjabi, the question remains: were there other triggers apart from retroflex plosives or other targets apart from dental plosives? The statistics in Table 18 and the evidence from historical-comparative data clearly indicate that other coronals have not participated in retroflex consonant harmony, either as triggers or targets of assimilation. The examples in (39) demonstrate this point with respect the retroflex flaps /ɽ/ and /ɽʰ/.

(39) No retroflex consonant harmony with flaps /ɽ/ and /ɽʰ/ (data from Turner, 1962–1966)

		Panjabi	Prakrit	Sanskrit	CDIAL
a.	‘palm tree’	taːɽ	taːɖa-	*taːɖa-	5750
	‘to break’	toːɽnaː	toːɖaï	troːʈajati	6079
	‘trunk of body’	dʰaɽ	dʰaɖa	*dʰaɖa	6712
	‘robbery’	dʰaːɽaː	dʰaːɖiː-	dʰaːʈiː	6772
	‘dust’	dʰuːɽ	dʰuːli-	*dʰuːɖi, dʰuːli-	6835
	‘grinder tooth’	daːɽʰ	daːɖʰaː-	dãʂʈra-	6250
	‘beard’	daːɽʰiː	daːɖʰiaː	daːɖʰikaː	6250
b.	‘quarrel’	raːɽ	raːɖi-	raːʈi	10697
	‘heap of manure’	ruːɽiː	ruːɖʰi-	ruːɖʰi	10802

c.	'vein'	naːɽiː	ɳaːɖiː	naːɖiː	7047
	'cane'	naɽaː	ɳaɖa-	naɖa-	6936
d.	'penis'	lauɽaː	la(g)uɖa-	lakuʈa	10875
	'oppression'	loɽʰaː	loːɖʰa	loːʈʰa	11134

The examples in (39) demonstrate that retroflex flaps do not trigger retroflex consonant harmony, either in plosives (39)(a) or in other sonorants (39)(b)–(d) (cf. Jain, 1934, p. 89). Note that the MIA forms in (39)(c) contain initial retroflex nasals (i.e., /ɳ…ɖ/). This is not the product of consonant harmony. It is a result of the complete neutralization of all coronal nasals to retroflex /ɳ/ in Prakrit (Masica, 1991, p. 182).[30]

The lack of harmony between flaps and plosives in (39)(a) is particularly striking given that retroflex flaps derive historically from plosives via lenition. This suggests that the lenition of plosives must have preceded the development of retroflex consonant harmony. By the time that harmony began to take effect, the intervocalic (singleton) retroflex plosives of OIA and MIA were already realized as sonorant flaps. Thus, they were no longer sufficiently similar to the plosives to participate with them in consonant harmony. In all of the cases where harmony has taken place (cf. examples in (33)), the trigger can be traced to a non-initial retroflex plosive that was either geminate or part of a homorganic cluster such as /-ɳɖ-/ or /-ʂʈ-/ (or OIA r + dental plosive, which often yielded geminate retroflex plosives in MIA (Masica, 1991, p. 176)). These are precisely the environments where retroflex plosives were retained as plosives

[30] It is possible that the single coronal nasal phoneme in MIA Prakrit was actually dental [n] initially and retroflex [ɳ] elsewhere. See Masica (1991, p. 182). Cf. footnote 25.

and, thus, were able to trigger harmony in word-initial dental plosives. Where the retroflex plosives of OIA and MIA were intervocalic singletons they became sonorant flaps. As such they failed to trigger harmony in word-initial plosives (cf. (39)(a)) because consonant harmony is sensitive to the relative similarity of triggers and targets.

Other coronal sonorants also failed to participate in harmony including the nasals /ɳ/ and /n/, as shown in (40).

(40) No retroflex consonant harmony with nasals /ɳ/ and /n/ (data from Turner, 1962–1966)

		Panjabi	Prakrit	Sanskrit	CDIAL
a.	'tune'	taːɳ	taːɳa-	taːna	5761
	'to stretch'	taːɳnaː	taːɳia-	taːnajati	5762
	'teat, udder'	tʰaɳ	tʰaɳa-	stana	13666
	'police station'	tʰaːɳaː, ʈʰaːɳaː	tʰaːɳa-, ʈʰaːɳa-	stʰaːna	13753
	'gift, charity'	daːɳ	daːɳa-	daːna	6265
	'incense burning'	dʰuːɳiː	dʰuːvaɳa-	dʰuːpana	6848
	'nape of neck'	dʰauɳ	dʰamaɳi-	dʰamani	6733
b.	'husband's sister'	naɳad, naɳaːn	ɳaɳãdaː-	nanaːndr̩	6946
	'strap'	niaːɳaː	ɳidaːɳa	nidaːna	7196
	'salt'	nuːɳ	luːɳa-	lavaɳa	10978
	'tank'	nauɳ	nahaːpana- (Pāli)	snaːpana-	13790a
	'butter'	nauɳiː	ɳavaɳiːa	navaniːta	7003

c.	‘dwarf’	naːʈaː	—	*naʈʈa-	6935
	‘to flee’	naʈʈʰɳaː	naʈʈʰa- (Pāli)	naṣʈa-, naʃjati	7027
	‘fearless’	niɖɖar	—	nirdara	7339
	‘boy, youth’	naɖɖʰaː	—	*naɖɖʰa-	6935

The examples in (40) demonstrate that retroflex /ɳ/ does not trigger harmony in dental plosives (40)(a) or in the dental nasal /n/ (40)(b). Moreover, dental /n/ is never the target of retroflex harmony, regardless of whether it is followed by retroflex /ɳ/ (40)(b) or a retroflex plosive (40)(c). The word /tʰaːɳaː/ ~ /ʈʰaːɳaː/ ‘police station’ in (40)(a) appears to show variable harmony in the initial plosive triggered by the retroflex nasal in C_2 position. However, this variation is not the product of consonant harmony. Notice that the variation is not unique to Panjabi but can be traced to MIA /tʰaːɳa-/ ~ /ʈʰaːɳa-/ ‘place’, which in turn corresponds to OIA /stʰaːna/ with initial /stʰ-/. Variation between /tʰ/ ~ /ʈʰ/ was a common MIA reflex of OIA /stʰ/ sequences (e.g., Pk. /ʈʰaːpeːti/ ‘establishes’ < Skt. /stʰaːpajati/; Pk. /attʰaː-, aʈʈʰaː-/ ‘trust’ < Skt. /aːstʰaː/) (Masica, 1991, pp. 172, 177). Thus, the variable retroflexion on the initial plosive is entirely independent of any influence of the retroflex nasal in C_2 position. Some of the MIA forms in (40)(b) contain two retroflex nasals (i.e., /ɳ…ɳ/). Once again, this is not the product of consonant harmony but of the neutralization of all non-labial nasals to retroflex /ɳ/ in Prakrit (Masica, 1991, p. 182). Panjabi /nuːɳ/ ‘salt’ in (40)(b) may exhibit an idiosyncratic case of nasal manner harmony (i.e., /n…ɳ/ < /l…ɳ/).[31] Even under these conditions retroflex assimilation has not occurred (i.e., */ɳ…ɳ/ < /l…ɳ/).

---

[31] Turner (1962–1966) has /nuːɳ/ ‘salt’ but Goswami (2000) has /luːɳ/ ‘salt’. This may reflect dialectal variation.

With respect to directionality, the situation in Panjabi is comparable to Dravidian (§3.1.4): the historical-comparative evidence clearly supports the application of regressive retroflex assimilation (T-Ṭ → Ṭ-Ṭ), which explains the avoidance of T-Ṭ configurations, but sheds no light on the absence of Ṭ-T configurations. As in Dravidian, the absence of Ṭ-T configurations is both expected and unexpected. It is expected because Old Indo-Aryan prohibited word-initial retroflexes, but it is unexpected given that Panjabi has developed word-initial retroflex plosives independent of consonant harmony. Synchronically, retroflex plosives occur word-initially before just about any segment class apart from dental plosives, including labial plosives (/ɖabbaː/ 'tin box'), velar plosives (e.g., /ʈuk-ɳaː/ 'to cut'), and anterior coronal sibilants and sonorants (e.g., /ʈass/ 'glamour'; /ɖãn/ 'fine, penalty'). The absence of Ṭ-T configurations can be explained in terms of progressive retroflex assimilation (Ṭ-T → Ṭ-Ṭ) or regressive dental assimilation (Ṭ-T → T-T). However, there is no historical-comparative evidence to support either of these patterns of assimilation. Rather, it seems that Panjabi lacked Ṭ-T configurations historically and simply failed to develop them. The failure to develop Ṭ-T configurations may reflect a principled co-occurrence restriction, but we cannot conclude anything about directionality concerning this restriction.

In sum, historical-comparative evidence indicates that retroflex consonant harmony has applied between coronal plosives in Panjabi, but not between plosives and sonorants. Moreover, all of the evidence indicates regressive retroflex assimilation; there is no indication of progressive retroflex assimilation or regressive dental assimilation.

3.2.3.3 Laryngeal features

While retroflex consonant harmony is clearly sensitive to manner of articulation in Panjabi, laryngeal features play no role. Agreement for laryngeal features is neither a condition for retroflex assimilation nor a necessary output of it. This is evident in Table 19, which shows pairs of coronal plosives in Panjabi classified according to agreement or disagreement for retroflexion and laryngeal features (voicing and aspiration). Pairs that agree in retroflexion are isolated from all others, including those that disagree in retroflexion and those that agree in non-retroflexion, because they are the only ones that are potential products of consonant harmony. There is no evidence that dental-dental pairs agreeing in non-retroflexion are the product of consonant harmony. Rather, they are the residue of consonant harmony after disharmonic T-Ṭ sequences have become harmonic Ṭ-Ṭ via retroflex assimilation. Counts are based on pairs of coronal plosives in headwords containing word-initial $C_1V(N)C_2$ sequences.

Table 19 Panjabi coronal plosives in #$C_1V(N)C_2$ sequences classified according to agreement for retroflexion and laryngeal features (n = 153)

		Retroflexion	
Laryngeal		**Agree**	**Other**
Agree	O	45	37
	O/E	3.97	1.14
Disagree	O	33	38
	O/E	1.06	0.49

The figures in Table 19 indicate that sequences of two coronal plosives that agree in both retroflexion and laryngeal features are over-attested in Panjabi. However, sequences of two coronal plosives that agree in retroflexion but disagree in laryngeal features are not under-attested as we might expect if agreement for laryngeal features were a necessary condition or output of retroflex consonant harmony. Rather, such sequences occur more-or-less as expected, as they do in Dravidian languages with retroflex consonant harmony (cf. §3.1.6). Thus, there is no evidence that laryngeal features play any role in the pattern of assimilation.

This conclusion is supported by the historical-comparative data reviewed in the preceding section (§3.2.3.2). Examples of retroflex harmony without agreement for laryngeal features, whether voicing or aspiration, include: /ʈiɖɖaː/ 'grasshopper' < MIA /tiɖɖa-/; /ɖaʈʈaː/ 'plug' < MIA /-daʈʈa-/; and /ʈuʈʈʰnaː/ 'to be kind' < MIA /tuʈʈʰa-/.

In sum, Panjabi exhibits a pattern of retroflex consonant harmony identical to that of Malto and other Dravidian languages: (i) it is root-internal; (ii) it is regressive (or possibly bidirectional); (iii) it applies only to pairs of coronal plosives, not to combinations of plosives and sonorants; (iv) it is not blocked by intervening segments; and (v) it does not entail agreement for laryngeal features.

3.2.4 The scope of retroflex consonant harmony in NIA

The preceding section reviewed evidence that Panjabi, a NIA language of the Central zone, exhibits the same pattern of retroflex consonant harmony as that found in Dravidian languages (reviewed in §3.1). A survey of Indo-Aryan languages reveals that Panjabi is not exceptional in this respect; it is representative of the vast majority of NIA languages of the Central, Northern and Northwestern zones (including the Dardic group, which is discussed independently in §3.3 below), and to some extent those of the Eastern zone, but not those of the Southern zone or the Sinhalese-Maldivian group.

In order to determine the scope of retroflex consonant harmony in Indo-Aryan, a study was conducted based on counts of headwords in dictionaries or other vocabulary sources for selected NIA languages. At least two languages were chosen to represent each geographic zone: 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 finally Sinhala and Dhivehi for the Sinhalese-Maldivian group.[32] For each language, counts were made of word-initial $\#C_1V(N)C_2$ sequences in which C_1 and C_2 are coronal plosives or retroflex sonorants and N is a homorganic nasal.[33] The results are presented in Table 20 and Table 21 following the convention of Pozdniakov & Segerer (2007) introduced in §3.1.3.

Table 20 $\#C_1V(N)C_2$ sequences in Indo-Aryan languages of the Northwestern, Northern, Central and Eastern zones (presented in that order).

Indus Kohistani (Zoller, 2005; n = 150)

	T	Ṭ	Ṛ
T	+ +	− −	+
Ṭ	− −	+ +	− −

Sindhi (Turner, 1969; n = 106)

	T	Ṭ	Ṛ
T	+	− −	
Ṭ	− −	+ +	− −

[32] The only zone not represented in the study is the East-Central zone, which consists of only five languages at most (Lewis, 2009). Data from two East-Central languages were examined: Awadhi (Turner, 1969) and Bagheli (Laiju Ek, p.c.). However, in each case the data was not sufficient for a statistical analysis. Thus, they are not included here. Cf. footnote 34 of this chapter.

[33] For each language the class of retroflex sonorants represented by Ṛ includes the following: Indus Kohistani /ɽ/; Sindhi /ɳ, ɽ, ɽ^h/; Kumauni /ɳ/ and [ɽ]; Nepali [ɳ, ɽ, ɽ^h]; Panjabi /ɳ, ɽ, ɽ^h/; Hindi /ɽ, ɽ^h/; Bangla /ɽ/; Oriya /ɳ, ɭ/ and [ɽ, ɽ^h]; Marathi /ɳ, ɭ/; Konkani /ɳ, ɭ/; Sinhala /ɳ, ɭ/; Dhivehi /ř, ɭ/. Segments shown here in square brackets [ ] may not be phonemic, though they are distinguished in the data source. Dhivehi /ř/ may be better classified as a retroflex sibilant [ʂ] rather than a sonorant. Either way the results are the same since the language does not exhibit retroflex consonant harmony of any kind.

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 = 145)

	T	Ṭ	Ṛ
T	+	–	
Ṭ	– –	+ +	

All of the languages in Table 20 show a tendency toward retroflex consonant harmony between plosives. In each case, sequences of plosives that agree in retroflexion are over-attested (i.e., Ṭ-Ṭ) while those that disagree for retroflexion are under-attested (Ṭ-T, T-Ṭ).

Exceptions to retroflex consonant harmony between plosives are mostly instances of T-Ṭ that fall into one of three categories: they are either (i) the product of variation between harmonic and disharmonic forms of the same word (T-Ṭ ~ Ṭ-Ṭ); (ii) unassimilated Sanskrit loanwords (i.e., *tatsamas*), which preserve archaic OIA T-Ṭ sequences; or (iii) morphologically complex words in which the two plosives are separated by an intervening morpheme boundary. For example, of 31 observed instances of initial T-Ṭ sequences in McGregor's (1993) Hindi dictionary, 16 are identified as direct Sanskrit loans and most others are either derived from these or exhibit either variation or morphological complexity. Some representative examples are listed in (41). If these categories are omitted the pattern of harmony approaches categorical status between plosives in most of the languages.

(41) Exceptions to retroflex consonant harmony in McGregor's (1993) Hindi dictionary

a. Variation between disharmonic and harmonic forms of the same word (T-Ṭ ~ Ṭ-Ṭ)

daʈnaː	~ ɖaʈnaː	'to be fixed, firm'
daɳɖaː	~ ɖaɳɖaː	'stick, pole'

b. Sanskrit loanwords (*tatsamas*)

taʈ	'bank (of river)'
taɳɖul	'grain (esp. rice)'
tuɳɖ	'snout, beak'
daɳɖ	'stick, staff'

c. Morphologically complex words

diʈʰvan 'waking of Viṣṇu' cf. devʈʰan < dev 'god' + uʈʰaːn 'rise'

With only one exception, all of the languages in Table 20 also show a lack of harmony between plosives and retroflex sonorants. In most cases retroflex sonorants occur either as expected with each class of coronal plosives, or disharmonic T-Ṛ sequences are preferred. The only exception to this rule is Nepali, which appears to prefer harmonic Ṭ-Ṛ sequences.[34] There are at least three possible explanations for the Nepali data: either (i) similarity is evaluated only

[34] Limited data for the East-Central Indo-Aryan languages, Awadhi (Turner, 1969) and Bagheli (Laiju Ek, p.c.), suggests that they too may exhibit a pattern of retroflex consonant harmony like that of Nepali with agreement for retroflexion not only between pairs of plosives but also between plosives and sonorants. Unfortunately, the data available for these languages is not sufficient to reach any reliable conclusions. Cf. footnote 32 of this chapter.

at a phonological level between contrastive categories; or (ii) similarity is evaluated over phonetic categories but consonant harmony preceded lenition diachronically in Nepali; or (iii) retroflex consonant harmony is not sensitive to similarity in Nepali and, therefore, applies to pairs of plosives and sonorants. Options (i) and (ii) are elaborated in the following paragraphs.

The Nepali data can be explained if similarity is evaluated only over phonemic representations. Of all the NIA languages included in the study, Nepali is the only one for which the entire class of retroflex sonorants is non-phonemic. Data for the present analysis was drawn from Turner's (1931) Nepali dictionary. The orthography and Roman transliteration in Turner's dictionary distinguish ⟨ɽ, ɽʰ⟩ from ⟨ɖ, ɖʰ⟩ though by all accounts the flaps are allophonic variants of the plosive (Acharya J. , 1991; Riccardi, 2003; Khatiwada, 2009). The complementary distribution of the two can be seen in Turner's data itself where the flaps do not occur initially and the plosives occur in non-initial position only as geminates or following a nasal. The dictionary also includes some cases of retroflex ⟨ɳ⟩ in C_2 position (i.e., without a following retroflex plosive). These are all identified via cross-reference as variants of forms with ⟨Ṽɽ⟩ (e.g., ʈaːɳ ~ ʈãːɽ 'platform'; ɖaːɳi ~ ɖãːɽi 'pole, staff') in which the nasal corresponds to a retroflex flap (with preceding nasalized vowel), which in turn is an allophone of the retroflex plosive /ɖ/. Thus, all instances of retroflex sonorants in C_2 position, whether ⟨ɽ, ɽʰ⟩ or ⟨ɳ⟩, can be traced to underlying retroflex plosives /ɖ, ɖʰ/. In a sense, then, it is possible to say that retroflex consonant harmony *does* hold only between plosives in Nepali but only if we assume that it applies to contrastive categories at a phonological level, and not to purely phonetic categories. In other words, the relevant class of participating segments may be defined as the class of phonemic plosives, which includes the allophonic variants [ɽ, ɽʰ] and [ɳ].

Alternatively, the pattern in Nepali can be explained in terms of phonetic similarity if we assume that consonant harmony applied between plosives before the lenition of retroflex plosives to sonorants in selected environments. If this is the case, then the relative diachronic sequencing of harmony and lenition in Nepali is different than that of other NIA languages where the absence of harmony between plosives and sonorants implies just the opposite ordering (cf. Panjabi in §3.2.3). At present, the most appropriate explanation for the Nepali data remains unclear and we must conclude that Nepali offers a potential, but not necessary, exception to the rule that retroflex consonant harmony in South Asian languages is sensitive to the similarity of participating segments. Further research may shed more light on the issue.

The Northwestern languages in Table 20 have some unique properties that deserve mention. Among the languages surveyed, Sindhi is unique in distinguishing a series of implosive stops. The dental vs. retroflex contrast is neutralized in the implosive series. As we might expect under these circumstances, the implosives do not participate in retroflex consonant harmony. In Turner's (1969) word list they co-occur freely with both dental and retroflex plosives.

Like most Northwestern languages of the Dardic sub-group, Indus Kohistani has contrastive retroflex sibilants (in addition to plosives and sonorants). The retroflex sibilants of Dardic languages do not exhibit harmony in relation to plosives or sonorants but do exhibit harmony in relation to other sibilants. Details concerning retroflex consonant harmony in the Dardic languages are examined independently in §3.3, below.

The pattern of retroflex consonant harmony between plosives appears to be strongest in languages of the Northwest, Northern and Central zones and somewhat weaker in those of the

Eastern zone. The two eastern languages, Oriya and Bangla, exhibit mixed systems in which a relatively large number of disharmonic T-Ṭ sequences are preserved alongside harmonic Ṭ-Ṭ sequences. Although T-Ṭ sequences occur less frequently than expected in both languages, they are not under-attested to the same degree that they are in the others. This is especially true of Oriya, which is the only language in Table 20 in which disharmonic T-Ṭ sequences are represented by a single " – " sign. In this respect, Oriya (and to a lesser degree, Bangla) bears a resemblance to Dravidian Gondi (cf. §3.1.3).[35]

While retroflex consonant harmony is strongest in NIA languages of the Northwest, Northern and Central zones and weakest in those of the Eastern zone, it is absent altogether in those of the Southern zone and the Sinhalese-Maldivian group, as shown in Table 21. Like OIA, the languages of the Southern zone and Sinhalese-Maldivian group generally avoid word-initial retroflex consonants (*Ṭ-T, *Ṭ-Ṭ) and preserve disharmonic T-Ṭ sequences intact.

Table 21 Observed counts for #$C_1V(N)C_2$ sequences in Indo-Aryan languages of the Southern zone and Sinhalese-Maldivian group

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

35 Eastern dialects of Gondi are in contact with Oriya in south-western Orissa (Lincoln, 1969; Steever, 1998a).

In languages that lack retroflex harmony and avoid word-initial retroflexes the few exceptional Ṭ-Ṭ and Ṭ-Ṛ forms are typically the product of other independent factors such as reduplication (often with an onomatopoeic function) or loanword adaptation. For example, the four exceptional Ṭ-Ṭ forms in Reynold's (2003) Dhivehi dictionary are all cases of reduplication (e.g., /(ham-)ʈoːʈoː/ 'old and wrinkled (skin)'; /ɖinɖin/ 'the female cuckoo'). Two other exceptions (not included in the count) are English loanwords in which coronal obstruents are adapted as retroflex (e.g., /ʈenʈu/ 'temporary cricket pavilion', cf. Eng. *tent*; /ʈeːʈaru/ 'theatre show', cf. Eng. *theatre*). In native vocabulary inherited from OIA, disharmonic T-Ṭ sequences are retained (e.g., /tuⁿɖu/ 'point', cf. Skt. /tuɳɖa/ 'beak'; /daⁿɖi/ 'stick', cf. Skt. /daɳɖa/ 'stick').

In sum, retroflex consonant harmony holds for most NIA languages of the Northwestern, Northern, Central and (to a lesser extent) Eastern zones. It does not hold for NIA languages of the Southern zone or for those of the Sinhalese-Maldivian group. Nor did it hold for OIA or MIA, though the early stages of retroflex consonant harmony are evident in the form of dialectal variation in MIA Prakrit. The pattern of retroflex consonant harmony in Indo-Aryan is virtually identical to that of Dravidian, with the possible exception of Nepali, which may violate the generalization that retroflex consonant harmony never holds between plosives and sonorants in South Asian languages.

3.3 Dardic & Burushaski

Although they are classified as Indo-Aryan languages of the Northwestern zone, the Dardic languages of northern Pakistan warrant independent discussion because they exhibit some unique properties. The Dardic languages are unique among NIA languages for preserving, not only the OIA contrast between dental and retroflex plosives (e.g., /t, ʈ/), but also the OIA contrast between dental, retroflex and palatal fricatives (i.e., /s, ʂ, ʃ/). Moreover, most of them have extended the three-way contrast to coronal affricates so that they now distinguish dental, retroflex and palatal affricates (i.e., /ʦ, ʈʂ, ʧ/). Thus, most Dardic languages exhibit contrastive retroflexion across three manners of articulation within the class of coronal obstruents: plosives, affricates and fricatives.

Contrastive retroflexion in both plosives and sibilants is a typologically rare phenomenon. To date, all previously reported cases of retroflex consonant harmony between obstruents occur in languages where retroflexion is contrastive only among plosives or sibilants, but not both (Arsenault & Kochetov, 2011). Thus, similarity effects are limited to cases where retroflexion is contrastive in plosives and sonorants, as in the Dravidian and Indo-Aryan languages reviewed in the preceding sections. The Dardic languages exhibit a striking and previously unattested similarity effect: retroflex consonant harmony in Dardic applies only to pairs of plosives or pairs of sibilants but not to mixed pairs of plosives and sibilants.

The following sub-sections present case studies of retroflex consonant harmony in two Dardic languages, each with slightly different properties: Indus Kohistani (§3.3.1) and Kalasha (§3.3.2). The scope of retroflex consonant harmony within the Dardic group is explored in §3.3.3. Finally, the isolate Burushaski is briefly discussed in §3.3.5 because of its geographic proximity and typological relation to the Dardic languages.

3.3.1 Indus Kohistani

Indus Kohistani is an Indo-Aryan language of the Dardic sub-group spoken in northern Pakistan. As shown in (42), Indus Kohistani exhibits contrastive retroflexion across three manners of articulation in its coronal obstruent system: plosives, affricates and fricatives. In addition it distinguishes at least one retroflex sonorant, the flap /ɽ/.[36] There is a clear contrast for manner between voiceless affricates and fricatives but not between voiced affricates and fricatives. The voiced palatal sibilants represented as /ʒ, ʒ^h/ tend to be pronounced as affricates word-initially and fricatives elsewhere (Zoller, 2005, p. 34; Hallberg & Hallberg, 1999, pp. 11, footnote 14). There is also a tendency to devoice voiced consonants and aspirate voiceless stops in word-final position (Zoller, 2005, p. 37). Thus, aspiration is contrastive word-initially but some instances of final aspiration may not be phonemic.

(42) Consonant phonemes of Indus Kohistani (Zoller, 2005; Hallberg & Hallberg, 1999)

LAB	DEN	RET	PAL	VEL	UVL	GLOT
p	t	ʈ		k	(q)	
p^h	t^h	ʈ^h		k^h		
b	d	ɖ		g		
b^h	d^h	ɖ^h		g^h		
	ts	ʈʂ	tʃ			
	ts^h	ʈʂ^h	tʃ^h			
(f)	s	ʂ	ʃ		(x)	h
	z	ʐ	ʒ		(ɣ)	
	z^h	ʐ^h	ʒ^h			
m	n					
	l					
	r	ɽ				
w			j			

[36] Indus Kohistani may also have a retroflex nasal phoneme /ɳ/ but according to Zoller (2005, p. 35) it is mostly realized as [ɽ] plus nasalization of the preceding vowel. Zoller's dictionary, which is the principle source of data for the present study, transcribes the actual pronunciation (i.e., one finds [Ṽɽ], not [Vɳ]). Thus, /ɳ/ is not counted independently in the present study (cf. Hallberg & Hallberg 1999, p. 18, footnote 24).

All retroflex obstruents occur in word-initial and non-initial positions. Only the retroflex sonorant /ɽ/ is banned word-initially. Word-initial retroflex obstruents are not necessarily the product of consonant harmony; they occur in roots containing labial and dorsal obstruents (e.g., /ʈikáṽ/ 'to dip'; /ɖʌbà^h/ 'a bundle'; /ʈʂùːk^h/ 'embroidery'; /ʂáːk^h/ 'neck'; /ʐʌ̀b/ 'a kind of long grass'; etc.).[37] Nevertheless, the present study demonstrates that the co-occurrence of coronal obstruents within roots is highly constrained by retroflex consonant harmony.

3.3.1.1 Synchronic coronal co-occurrence patterns in Indus Kohistani

In order to explore synchronic co-occurrence restrictions on coronal obstruents a study was conducted based on data from Zoller's (2005) dictionary of Indus Kohistani. The dictionary was searched for lexical entries containing word-initial $C_1V(N)C_2$ sequences in which C_1 and C_2 are coronal obstruents and N is a homorganic nasal. Rather than counting headwords in this case, an attempt was made to reduce the data to a more restrictive set of unique roots.[38] In order to achieve this, lexical entries were included or excluded from the data set based on the following criteria:

---

[37] Here and elsewhere, all data examples for Indus Kohistani are drawn from Zoller (2005). Zoller's transcription includes some phonetic details that may not be phonemic (2005, p. 34–35). I have adapted his transcription to IPA conventions, but otherwise I have preserved it without attempting to eliminate redundant elements. Phonetic details that may be redundant include: the vowel [ʌ], which can be regarded as an allophone of /a/; word-final devoicing of voiced segments (e.g., /adʌ̀ɖ̥/ 'thing'); most cases of word-final aspiration (e.g., /sù^h/ 'he'); and most final ultrashort vowels, which are represented as superscript vowels (e.g., /dít^{hi}/ 'given'). Accents on vowels represent rising (V́) and falling (V̀) pitch accents.

[38] The overall results of the Indus Kohistani case study are essentially the same whether counts are based on all headwords in the dictionary or the more restrictive set of unique roots. This finding validates the coarse-grained method of counting headwords adopted elsewhere in the dissertation. Counts based on roots are presented here simply because they were available. For comparison, counts based on headwords can be found in Appendix B.

- Neighbouring languages: Zoller's dictionary of Indus Kohistani includes lexical entries for words drawn from the neighbouring languages Gabār (a.k.a. Gowro) and Bhaṭīsē (a.k.a. Baṭēṛā). These were excluded on the grounds that they are generally regarded as distinct languages, not dialects of Indus Kohistani. Their lexical similarity with Indus Kohistani is only 61% and 58%, respectively (Lewis, 2009).
- Dialectal variants: Zoller's dictionary is based primarily on the Jijālī dialect of Indus Kohistani but also includes some data from the Šāṭōṭī dialect. Wherever variations of the same word were listed for both dialects only the Jijālī word was counted in order to avoid duplicates of a single root. A few words from Šāṭōṭī were retained but only when the Jijālī variant was not available. A small number of roots exhibit variation between harmonic and disharmonic forms. In these cases the harmonic variant was counted (i.e., a root was counted as harmonic if it was at least *potentially* so). For examples of this kind see (47) below.
- Derived and inflected variants were excluded in order to avoid duplicates of a single root. For instance, where masculine and feminine forms of a word were both listed, or nominal and adjectival forms, then only one representative of each was counted. Thus, the masculine adjective /tʌ̀tʰ/ 'hot' was counted but its feminine counterpart /tʌ̀jtʰⁱ/ and its derived nominal counterparts /tʌtìː/ 'heat' and /tʌtrúː/ 'hot ashes' were not.
- Morphologically complex words were excluded if an identifiable morpheme boundary occurred between C_1 and C_2. Among other things this criterion excluded a large number of reduplicated forms, most of them onomatopoetic (e.g., /ʧũ̀ː-ʧãː/ 'inarticulate sounds produced by a dumb person'; /ʈ͡ʂã́ː-ʈ͡ʂãː/ 'sound of panicking goats'; /ʐõ̀ː-ʐõː/ 'sound of a buzzing or humming insect', etc.).

- English loanwords were excluded but roots of Persio-Arabic origin were retained since they constitute a relatively substantial portion of the contemporary Indus Kohistani lexicon (at least 15% by Zoller's estimation (2005, p. 16)). A small number of roots that may be loanwords from other Indo-Aryan or non-Indo-Aryan languages of the area were also retained (e.g., Shina, Urdu, Burushaski, etc.).

Coronal obstruents were collapsed into five natural classes based on place and manner but ignoring laryngeal distinctions: dental plosives, retroflex plosives, dental sibilants, palatal sibilants, and retroflex sibilants. Each root was then classified as belonging to one of 15 logically possible combinations of C_1 and C_2, ignoring the relative order of consonants. For instance, words such as /ɖʰús/ 'bruised', /zàːʈʰ/ 'body hair, wool', /ʈèːʦʰ/ 'flint' and /ʦaʈáṽ/ 'to lick' were all classified as representing the combination of retroflex plosives with dental sibilants. The results are displayed in Table 22.

Table 22 Coronal obstruents in #C_1V(N)C_2 sequences in Indus Kohistani roots; observed counts and O/E ratios (n = 303)

C_1 \ C_2	t, tʰ, d, dʰ	ʈ, ʈʰ, ɖ, ɖʰ	ʦ, ʦʰ, s, z, zʰ	ʧ, ʧʰ, ʃ, ʒ, ʒʰ	ʈʂ, ʈʂʰ, ʂ, ʐ, ʐʰ
t, tʰ, d, dʰ	31 1.21	2 0.07	69 1.66	35 0.97	10 0.54
ʈ, ʈʰ, ɖ, ɖʰ		21 2.59	28 1.01	30 1.18	10 0.84
ʦ, ʦʰ, s, z, zʰ			15 1.01	9 0.37	1 0.07
ʧ, ʧʰ, ʃ, ʒ, ʒʰ				21 2.24	0 0.00
ʈʂ, ʈʂʰ, ʂ, ʐ, ʐʰ					21 (6.43)

Several observations can be made based on the results in Table 22. First, combinations of two coronal obstruents that belong to the same manner class and agree in place of articulation all occur at or (more often) above their expected frequency. This includes combinations of two plosives (both retroflex O/E = 2.59; both dental O/E = 1.21) or two sibilants (both retroflex O/E = 6.43; both palatal O/E = 2.24; both dental O/E = 1.01). Second, and more importantly, combinations of two coronal obstruents that belong to the same manner class but disagree in place of articulation are either categorically absent or substantially under-attested. This includes combinations of two plosives (dental/retroflex O/E = 0.07) or two sibilants (palatal/retroflex O/E = 0.00; dental/retroflex O/E = 0.07; dental/palatal O/E = 0.37). The four lowest O/E ratios in the table correspond to the four possible combinations of obstruents that belong to the same manner class but disagree for place of articulation.

Third, combinations of two coronal obstruents with different manners of articulation (i.e., plosives with sibilants of any kind) are relatively unconstrained with respect to place of articulation. Some plosive/sibilant combinations are over- or under-attested to a degree. Nevertheless, they lack the extreme polarization evident in those cells representing plosive/plosive and sibilant/sibilant combinations where over-attestedness of pairs agreeing for place of articulation is accompanied by categorical or near-categorical absence of pairs disagreeing for place. For instance, T-S/S-T combinations agreeing for dental place occur more frequently than expected (O/E = 1.66) but Ṭ-S/S-Ṭ combinations disagreeing for place are not under-attested (O/E = 1.01). Similarly, T-Ṣ/Ṣ-T combinations disagreeing for place are under-attested (O/E = 0.54) but Ṭ-Ṣ/Ṣ-Ṭ combinations agreeing for place are also slightly under-attested (O/E = 0.84), not over-attested as we might expect if harmony applied in these cases.

Thus, Indus Kohistani exhibits a pattern of coronal place harmony that is sensitive to the similarity of participating obstruents in terms of their manner of articulation. Coronal obstruents in a root must agree for place of articulation but only if they share the same manner. Within the sibilant class, where three places of articulation are distinguished, agreement for place appears to go beyond retroflex harmony. Not only are dental-retroflex and palatal-retroflex combinations avoided but also combinations of dental and palatal sibilants (O/E = 0.37). Agreement is nearly categorical for retroflexion in both the plosive and sibilant classes but agreement between palatal and dental sibilants is somewhat gradient; a number of exceptions do occur. Thus, the language clearly exhibits retroflex harmony but also approaches a three-way coronal place harmony among sibilants, at least as a statistical tendency.

Further details emerge when the class of sibilants is expanded into affricate and fricative sub-classes, as shown in Table 23. Recall that voiced palatal sibilants are realized as phonetic affricates initially and as fricatives elsewhere. For this reason initial /ʒ-, $ʒ^h$-/ are classified with palatal affricates in Table 23 while non-initial /-ʒ, -$ʒ^h$/ are classified with palatal fricatives.

Table 23 Coronal obstruents in $\#C_1V(N)C_2$ sequences in Indus Kohistani roots with sibilants expanded into affricate and fricative classes (n = 303)

C_1 / C_2	t, tʰ, d, dʰ	ʈ, ʈʰ, ɖ, ɖʰ	ʦ, ʦʰ	ʧ, ʧʰ, ʒ-, ʒʰ-	ʈʂ, ʈʂʰ	s, z, zʰ	ʃ, -ʒ, -ʒʰ	ʂ, ʐ, ʐʰ
t, tʰ, d, dʰ	31 1.21	2 0.07	5 0.76	13 0.71	5 0.73	64 1.83	22 1.24	5 0.43
ʈ, ʈʰ, ɖ, ɖʰ		21 2.59	9 1.78	22 1.58	2 (0.41)	19 0.84	8 0.70	8 1.15
ʦ, ʦʰ			1 (8.42)	0 (0.00)	0 (0.00)	4 (1.14)	0 (0.00)	0 (0.00)
ʧ, ʧʰ, ʒ-, ʒʰ-				3 (2.65)	0 (0.00)	6 0.61	9 1.71	0 (0.00)
ʈʂ, ʈʂʰ					2 (6.31)	1 (0.25)	0 (0.00)	10 (6.23)
s, z, zʰ						10 0.90	3 0.26	0 0.00
ʃ, -ʒ, -ʒʰ							9 (3.04)	0 (0.00)
ʂ, ʐ, ʐʰ								9 (6.68)

When sibilants are expanded into affricate and fricative classes the observed and expected values for each C_1-C_2 combination drops resulting in exaggerated O/E values for most cells representing sibilant-sibilant pairs in Table 23. As a result, any conclusions drawn from the data must be tentative at best. Nevertheless, it is possible to make some observations. First, the three roots with disharmonic dental/palatal sequences involving two fricatives are all roots of Persio-Arabic origin. Within the stock of native Indo-Aryan roots all exceptions to sibilant harmony involve affricate/fricative pairs while sibilants with identical manner (i.e., affricate/affricate and fricative/fricative pairs) exhibit a more categorical pattern of place agreement. This suggests that, even among sibilants, harmony is enforced more strictly between segments with identical manner of articulation than between those with slightly different manners of articulation.[39]

Second, all exceptions to sibilant harmony involve dental/palatal and dental/retroflex pairs. There are no exceptions in the data set involving palatal/retroflex pairs (whether affricates or fricatives). This suggests that harmony among posterior sibilants (i.e., palatal and retroflex) might be enforced more strictly than harmony between anterior (i.e., dental) and posterior sibilants.

Third, palatal affricates are less constrained in relation to dental fricatives (O/E = 0.61) than retroflex fricatives (O/E = 0.00). Recall that palatal affricates originated as part of the OIA stop system but fail to participate in consonant harmony with other stops (i.e., dental and retroflex plosives) presumably because of their status as sibilant affricates. However, they also fail to participate in harmony with dental fricatives and this may be due in part to their status as non-continuant stops. Thus they appear to fall in the gap between the stop and sibilant classes. If they participate in consonant harmony at all it is only with retroflex sibilants and possibly with the class of affricates.

---

[39] If all voiced sibilants are classified as phonological affricates (on the grounds that they developed from affricates via lenition) then this generalization is strengthened further. Under this analysis, two of the three Persio-Arabic roots with disharmonic dental/palatal sequences involving two fricatives would be re-classified as affricate/fricative pairs.

3.3.1.2 Historical-comparative evidence of consonant harmony

Historical and comparative data support the observations concerning retroflex consonant harmony in Indus Kohistani and reveal some other properties of the pattern. Consider the examples in (43) and (44).

(43) Evidence of retroflex consonant harmony between plosives in Indus Kohistani

	Ind. Kohistani	MIA	OIA	CDIAL
'small horse'	ʈʌʈúː	—	*taʈʈu[40]	5440
'small rug'	ʈàːʈʰ	ʈaʈʈiː-, taʈʈiː-	*traʈʈa	5990
'clever, cheerful'	ʈʰàːʈʰ	taʈʈʰa-	taʂ̥ʈa-	5743
'span of hand'	ɖíːʈʰi	*diʈʈʰi-	diʂ̥ʈi	6343
'trustworthy'	ɖaːɖíː	daɖʰa-	daːrɖʰja-, dr̥ɖʰa-	6508
'stick'	ɖáːɳɖ̊	ɖãɖa-, dãɖa-	daɳɖa	6128

[40] Zoller (2005) identifies this OIA cognate as */ʈaʈʈu/ (not */taʈʈu/). His source is Turner (1966), who reconstructs it as either */ʈaʈʈu/ or */taʈʈu/. I assume that the disharmonic form */taʈʈu/ is original on the grounds that word-initial retroflex consonants were rare or non-existent at one time in OIA (Masica, 1991, p. 157). The disharmonic form is attested in at least one other New Indo-Aryan language, i.e., Marathi /taʈʈuː/ 'pony'.

(44) Evidence of retroflex consonant harmony between sibilants in Indus Kohistani [41]

		Ind. Kohistani		OIA	CDIAL
a.	‘spotted’	ʈʂìʈʂʰ	*ʧiʈʂ- (ʈʂ < tr)	ʧitra-	4803
	‘son-in-law’	ʐʌmʈʂú:	Sh. ʒamʈʂo: (ʈʂ < tr)	*ʤa:ma:traka-, ʤa:ma:tṛ	5198
	‘to learn’	ʈʂʰiʈʂáṽ	Sh. siʈʂ (ʈʂ < kʂ)	ʃikʂate	12430
	‘sudden pain’	ʈʂʌ̀ʂ		ʧaʂati	4727a
	‘to suck’	ʈʂo:ʂáṽ	Sh. ʧu:ʂ	ʧu:ʂati	4898
	‘hill’	ʈʂʰì:ʐ	Sh. ʧʰi:ʂ	—	
b.	‘decent, fine’	ʂù:ʂ		su:kʂma	13546
	‘straight’	ʂũ̀:ʂ	Sh. sũ:ʈʂok	*su:ŋkʂa	13548
	‘head’	ʂìʂ		ʃi:rʂa	12497
	‘left over’	ʂiʂã̀:		ʃe:ʂa	12611
	‘light wind’	ʂĩ̀:ʂʌṽ hʌvá:		ʃuʂila-	12547

In each of the examples in (43) and (44) retroflex consonant harmony exhibits the following properties: (i) it is regressive; (ii) it is triggered by retroflex segments; (iii) it targets dental or palatal segments; and (iv) it holds only between obstruents of the same manner class, whether plosives or sibilants. The examples in (43) provide evidence of retroflex harmony between plosives much like the pattern seen in Panjabi (cf. (33)) and other South Asian

41 MIA cognates from Prakrit and Pāli are not provided in (44) because these languages did not distinguish retroflex sibilants of any kind. In place of MIA, some cognates are provided from Shina (Sh.), another Dardic language that appears to preserve some disharmonic forms involving retroflex sibilants (cf. discussion in 3.3.3).

languages. Those in (44) provide evidence of harmony between sibilants. In (44)(a) initial retroflex sibilants of Indus Kohistani (/ʈʂ/ or /ʐ/) correspond to OIA palatal sibilants (/ʧ/, /ʤ/ or /ʃ/) that were followed by retroflex sibilants (/ʈʂ/ or /ʂ/) at a distance. Old Indo-Aryan did not have retroflex affricates to serve as triggers but they developed in Indus Kohistani and other Dardic languages from OIA /kʂ/ or /Cr/ sequences (e.g., Indus Kohistani /ʈʂʰàːr/ 'waterfall' < OIA /kʂara-/ 'melting away'; /ʈʂàː/ 'three' < OIA /trajaḥ/ 'three'). Wherever this occurred in C_2 position, the resulting retroflex affricate triggered retroflex harmony in any preceding sibilants. In (44)(b), initial retroflex fricatives correspond to OIA dental or palatal fricatives (either /s/ or /ʃ/) that were followed by retroflex sibilants at a distance.[42]

Retroflex consonant harmony did not apply to pairs of obstruents if one was a sibilant and the other a plosive, as demonstrated by the examples in (45).[43] Nor did it apply between plosives if the retroflex plosive in C_2 position developed into a sonorant flap through lenition, as demonstrated in (46)(a).

---

[42] In some of the roots in (44)(b), Indus Kohistani /ʂ/ in C_2 position corresponds to OIA /kʂ/ (e.g. /ʂùːʂ/ 'decent, fine' < OIA /suːkʂma/). In these cases it is not clear whether harmony was triggered by the OIA fricative /ʂ/ or by a retroflex affricate /ʈʂ/, which is the expected reflex of OIA /kʂ/ in Indus Kohistani. It is also unclear whether the absence of the expected /ʈʂ/ in Indus Kohistani is in any way connected to the harmony (i.e., whether retroflex harmony between sibilants can also result in agreement for manner along the affricate/fricative dimension) or whether it is simply the result of lenition, which is also typical of affricates the language.

[43] The words /tʰʌʂ/ 'slick, slippery' and /ʈʰʌʂ/ 'a slip, slide' appear to be derivatives of a common root. If so, then they may constitute an exception to the generalization that retroflex harmony does not hold between plosives and sibilants. The exceptional nature of this example is evident when compared to the other examples in (45) and to the statistics in Table 22.

(45) No retroflex consonant harmony between obstruents with different manners

	Ind. Kohistani	MIA	OIA	CDIAL
‘to carve’	taʈʂʰáṽ	(ʈʂʰ < kʂ)	takʂati	5620
‘(at the) right’	daʈʂʰóː, daʂȍ̃ː	(ʈʂʰ < kʂ)	dakʂiɳa	6119
‘sin’	dùːʂ		doːʂa	6587
‘a whistle’	siʈìː		*siːʈʈa	13427
‘rich’	sìːʈʰ	seʈʈʰi-, siʈʈʰi-	ʃreːʂʈʰin	12726
‘hair bun’	ʧȍ̃ːʈʰ	ʧoʈʈiː-, ʧuːɖaː-	*ʧoːɳɖa-, ʧuːɖa-	4883
‘name of a month’	ʒèːʈʰ	ʤeʈʈʰa-	ʤjaiʂʈʰa	5293
‘a bump’	ʃòʈʰ		*ʃoːʈʈʰa	12513

(46) No retroflex consonant harmony between plosives and retroflex flaps

		Ind. Kohistani	MIA	OIA	CDIAL
a.	‘a blow, knock’	tʌɽʌ̀qʰ	—	*taɖati	5632
	‘to weigh’	dʰʌ́ɽ karʌ́ṽ	—	dʰaʈa-	6706
	‘body’	dʰʌ́ɽ	dʰaɖa-	*dʰaɖa-	6712
	‘cattle raid’	dʰaːɽàʰ	dʰaːɖiː-	dʰaːʈiː	6772
	‘dust’	dʰúɽ	dʰuːlĭː	*dʰuːɖi, dʰuːli	6835
b.	‘cliff’	tʌ̀ɽaː ~ ʈʌ̀ɽaː	taɖa-	taʈa-	5629

Once again, the absence of harmony in (46)(a) is striking given that the retroflex flaps of Indus Kohistani derive from OIA retroflex plosives. In each case the flap corresponds to an OIA intervocalic singleton plosive that was subject to lenition. The sonorant flap that resulted from lenition failed to trigger harmony in dental plosives. In all of the cases where harmony has applied between plosives, the trigger can be traced to an OIA retroflex plosive that was either geminate or part of a consonant cluster that reinforced and preserved its plosive manner (cf. examples in (43)). The example /tʌ̀ɽaː ~ ʈʌ̀ɽaː/ in (46)(b) is exceptional in this respect. It is the only root in the data set to show variable harmony triggered by a retroflex flap.

Although the statistical analysis in §3.3.1.1 reveals a tendency toward a full three-way coronal place harmony, historical-comparative data supporting assimilation between palatal and dental sibilants is scarce. Most palatal-palatal pairs can be traced to OIA forms that already contained two palatals (e.g., /ʃĩʃʌṽ/ 'a poplar' < OIA /ʃĩʃapaː-/; CDIAL 12424). A possible case of palatal assimilation is /ʃòːʃaː/ 'decoration' if it derives from OIA /*suʃoːbʰa/ 'splendid' (CDIAL 13534) as suggested by Zoller (2005, p. 387). However, even this example is doubtful. Alternative sources include OIA /ʃoːʃutʃat/ 'shining brightly' (CDIAL 12642) or, more likely, Persio-Arabic /ʃaʻʃaʻa/ 'tail of a letter, cedilla' via colloquial Hindi-Urdu /ʃoʃaː/ 'tail (of an Arabic letter)' with the sense of 'decoration' or 'embellishment' (Steingass, 1892; Platts, 1884; McGregor, 1993). Thus, while retroflex consonant harmony among sibilants receives robust support from both statistical counts and historical-comparative data, the status of palatal harmony remains uncertain.

3.3.1.3 Exceptions to retroflex consonant harmony

Exceptions to retroflex consonant harmony in Indus Kohistani typically fall into one of two categories: (i) disharmonic forms with harmonic variants; or (ii) morphologically complex forms in which an identifiable morpheme boundary intervenes between C_1 and C_2. Examples of synchronic variation between harmonic and disharmonic forms of the same root are shown in (47).

(47) Synchronic variation between harmonic and disharmonic forms of a root

a. ʈʰóːʈʰ ~ tʰóːʈʰ ~ ʈʰòːto ‘fur shoes’

ɖíːʈʰⁱ ~ díʈʰ ‘span of hand’

b. ʂũ̀ːʂ ~ sũ̀ːʂ ‘straight’

ʈʂʌ̀ʂ ~ ʈʂʌ̀s ‘sudden fierce or stabbing pain’

ʐʰʌ̀ʂ ~ ʐʰʌ̀s ‘a shift, move, pull’

Some of the examples in (47) suggest the possibility of progressive retroflex harmony (e.g., /ʈʂʌ̀ʂ ~ ʈʂʌ̀s/). However, it is unclear whether these are the product of progressive harmony or other independent developments. For instance, /ʈʂʌ̀ʂ/ ‘sudden fierce or stabbing pain’ probably derives from OIA /ʧaʂati/ ‘hurts’ (CDIAL 4727a) via regressive retroflex assimilation. If so, then the disharmonic variant /ʈʂʌ̀s/ must be a later development involving the loss of harmony, possibly the result of interference from a competing root with overlapping semantics, such as /ʦàs/ ‘a pinch, a sudden pull’ from OIA */ʧassakk/ ‘throb, twitch, sudden pain’ (CDIAL 4730). The data available at present is not sufficient to resolve the issue; progressive harmony remains a possible, but doubtful, explanation for some cases of variation in (47).

Most other exceptions to retroflex consonant harmony involve morphological complexity, as shown in (48).

(48) No harmony across morpheme boundaries

a. ɖʌ̀j̃-tapxʌ̀j ‘spine’

b. ʂʌ̀-s — ‘this (OBL)’ — ← ʂũ̀ʰ ‘this’ (NOM)

ʂʌ̀-zeː — ‘here, at this place’ — ← ʂʌ̀-s ‘this (OBL)’ + zéː ‘place’

ʂú-ʃàːnãː — ‘similar’ — ← ʂũ̀ʰ ‘this’ + ʃaːn ‘resembling’

ʂa-z-ʌ́ṽ — ‘to be attached’ — ← ʂáṽ ‘to attach’ + passive -z-

ʐʰaː-zʌ́j — ‘brother's wife’ — ← ʐʰàː ‘brother’ + ?? [44]

The examples in (48) demonstrate that harmony does not extend across morpheme boundaries. They can also be explained if consonant harmony is purely regressive since each example contains a retroflex obstruent in C_1 position that fails to trigger harmony in a following obstruent of the same manner class. The data is consistent with either restriction and it is not clear whether only one is valid or whether both hold simultaneously.

3.3.1.4 Laryngeal features

Consonant harmony among obstruents in Indus Kohistani does not necessarily entail agreement for laryngeal features. Very few C_1-C_2 pairs of any kind show agreement for aspiration or breathy voice (regardless of whether cases of final aspiration are taken as phonemic or not). This is not surprising given that a form of Grassman’s law – a dissimilatory rule prohibiting sequences of two aspirates in successive syllables – operated in Sanskrit (MacEachern, 1997) and is still evident in Dardic languages such as Palula (Liljegren, 2008).

Leaving aspiration aside, the question remains whether consonant harmony entails voicing agreement. In order to explore this question all roots with initial $C_1V(N)C_2$ sequences

---

[44] The morpheme /-zʌj/ is probably cognate with OIA /ʤaːjaː/ ‘wife’ (CDIAL 5205).

containing two coronal obstruents were classified into one of two major categories: (i) roots that agree for retroflexion and manner (i.e., those that are potentially the product of retroflex consonant harmony) and (ii) all others – i.e., those that disagree for retroflexion and/or manner and those that agree in non-retroflexion.[45] Each of these categories was further sub-classified into (i) roots that agree in voicing (whether voiced or voiceless) and (ii) roots that disagree in voicing.[46] O/E ratios were calculated based on sums of observed and expected counts for each of the four possible sub-categories. The results are presented in Table 24.

Table 24 Coronal obstruents in $\#C_1V(N)C_2$ sequences in Indus Kohistani roots classified according to agreement for retroflexion, manner and voice (n = 303)

		Retroflexion & Manner	
Voice		Agree	Other
Agree	O	32	158
	O/E	4.83	0.98
Disagree	O	10	103
	O/E	(2.10)	0.79

The results in Table 24 reveal a preference for voicing agreement in roots that exhibit retroflex consonant harmony (O/E = 4.83) but one that is far from absolute. While harmonic

[45] For the purpose of these counts affricates and fricatives were counted as a single manner because retroflex harmony clearly operates between them. Thus, only two manner classes were distinguished: plosives and sibilants.

[46] Word-final phonetically devoiced obstruents (e.g., [d̥] in Zoller's transcription) were classified as voiced for the purpose of this analysis on the grounds that they are voiced phonemically and that their devoicing is conditioned by factors independent of consonant harmony.

roots that disagree in voicing are fewer and have a lower overall O/E value relative to those that agree in voicing, they are not under-attested (O/E = 2.10).[47] Moreover, historical-comparative evidence clearly indicates that voicing agreement has not necessarily accompanied retroflex consonant harmony (e.g., /ɖíːʈ^{hi} ~ díʈ^{h}/ 'span of hand' < OIA /diʂʈi/). In fact, most of the roots that exhibit both retroflex harmony and voicing agreement already had voicing agreement before retroflex harmony applied (e.g., /ɖaːɖíː/ 'trustworthy' < OIA /daːrɖ^{h}ja-/; /ʂìʂ/ 'head' < OIA /ʃiːrʂa/; cf. other examples in (43) and (44)).

Thus, laryngeal features play little or no role in retroflex consonant harmony in Indus Kohistani, either in the conditioning of harmony or in its output. While obstruents that agree for retroflexion and manner also tend to agree for voicing, the tendency is far from absolute and there is no evidence to suggest that it is the product of retroflex consonant harmony.

In sum, Indus Kohistani exhibits a pattern of retroflex consonant harmony with striking similarity effects. Harmony holds only between coronal obstruents of the same manner class, whether plosive or sibilant, but not between obstruents of different manner classes (i.e., plosives and sibilants). Nor does it hold between obstruents and sonorants. The following section presents a case study of Kalasha, another Dardic language with retroflex consonant harmony, but one with slightly different similarity effects.

---

[47] The O/E value of 2.10 for obstruent pairs that agree in retroflexion and manner but disagree in voicing is enclosed in parentheses in Table 24. This indicates that it is potentially exaggerated because it is based on an expected value that is less than 5.0. However, in this case the O/E value is probably not far from the mark because the expected value is 4.8, only just below the desired minimum of 5.0.

3.3.2 Kalasha

Kalasha is another Indo-Aryan language of the Dardic sub-group spoken in northern Pakistan. In a collection of notes on the historical development of Kalasha consonants, Morgenstierne (1973, p. 201) observed a few cases of "assimilation at a distance", most of them involving retroflexion. More recently, Arsenault & Kochetov (2009; 2011) have explored the topic of retroflex consonant harmony in Kalasha in greater detail. This section presents the results of that study.[48]

Like most other languages of the area, Kalasha has a rich inventory of coronal obstruents that includes retroflex plosives, affricates and fricatives. The consonant phonemes of Kalasha are listed in (49).

[48] Arsenault & Kochetov (2009; 2011) presented preliminary results of the Kalasha case study that were based only on counts of word-initial $\#C_1VC_2$ sequences in roots. The present discussion incorporates much subsequent work that examined a broader range of data. The overall results presented here are essentially the same but some of the particulars are not.

(49) Consonant phonemes of Kalasha (Trail & Cooper, 1999; Heegård & Mørch, 2004)

LAB	DEN	ALV	RET	PAL	VEL	GLOT
p	t		ʈ		k	
p^h	t^h		ʈ^h		k	
b	d		ɖ		g	
b^h	d^h		ɖ^h		g^h	
	ʦ		ʈʂ	ʧ		
	ʦ^h		ʈʂ^h	ʧ^h		
	ʣ		ɖʐ	ʤ		
				ʤ^h		
	s		ʂ	ʃ		h
	z		ʐ	ʒ		
m	n		(ɳ)		ŋ	
	ł	l				
		r	(ɽ)			
w				j		

All dialects of Kalasha distinguish retroflex plosives, affricates and fricatives from their dental counterparts and, in the case of sibilants, also from their palatal counterparts. Some dialects of Kalasha may also distinguish the retroflex sonorants /ɳ/ and /ɽ/ (Heegård & Mørch, 2004). Unfortunately, neither of these phonemes is distinguished in Trail & Cooper (1999), which is the principal source of data for the present study. Thus, their co-occurrence with other consonants is not explored here. Like Indus Kohistani, the retroflex obstruents of Kalasha can occur word-initially or non-initially. Their co-occurrence with non-coronal obstruents is unrestricted (e.g., /ʈak/ 'stingy'; /kaʈ/ 'board'; /ʈʂap/ 'completely through'; /paʈʂ/ 'feather'; /ʂup-ik/ 'to attack'; /p^{h}uʂ/ 'breath'; etc.), but their co-occurrence with other coronal obstruents is constrained by retroflex consonant harmony.

3.3.2.1 Synchronic coronal co-occurrence patterns in Kalasha

In order to explore synchronic co-occurrence restrictions on coronal obstruents in Kalasha a study was made of lexical roots in the language. Data for the study was drawn from an electronic lexical database compiled by Ron Trail and Greg Cooper. The database is an expanded version of the one used by Trail & Cooper for their (1999) Kalasha dictionary. The data was searched for all instances of $C_1 \ldots C_2$ in which C_1 and C_2 are both coronal obstruents separated minimally by an intervening vowel and potentially by any number of intervening vowels or consonants. The resulting list was reduced to a more restrictive set – one that approximates the set of unique roots in the language – by excluding items that were either (i) derived or inflected forms of another root already included in the count or (ii) morphologically complex forms in which an identifiable morpheme boundary occurs between C_1 and C_2.[49] A small set of English loanwords was also excluded but all other roots were retained, including those of Persio-Arabic origin and a few potential loanwords from neighbouring Indo-Aryan languages such as Khowar and Urdu. These criteria yielded a set of 766 roots.

Coronal obstruents were collapsed into eight natural classes based on place and manner regardless of laryngeal features: dental plosives, retroflex plosives, dental affricates, palatal affricates, retroflex affricates, dental fricatives, palatal fricatives and retroflex fricatives. Each

---

[49] It was not possible to determine the morphological structure of each and every word with absolute certainty. Thus, it is possible (even likely) that some items retained in the set of unique roots are in fact morphologically complex items that should have been excluded. The number of such items is probably quite small and it is doubtful that their exclusion from the data set would alter the overall results of the study. This is supported by the fact that the overall results are the same even when all derived, inflected and/or morphologically complex items are retained (i.e., the count of all headwords meeting the search criteria).

root was then classified as belonging to one of 36 logically possible combinations of C_1 and C_2 ignoring the relative order of consonants. The results are displayed in Table 25.

Table 25 Coronal obstruents in #C_1…C_2 sequences in Kalasha roots; observed counts with O/E ratios (n = 766)

C_1 / C_2	t, tʰ, d, dʰ	ʈ, ʈʰ, ɖ, ɖʰ	ʦ, ʦʰ, ʣ	ʧ, ʧʰ, ʤ, ʤʰ	ʈʂ, ʈʂʰ, ɖʐ	s, z	ʃ, ʒ	ʂ, ʐ
t, tʰ, d, dʰ	70 0.95	7 0.12	12 1.41	67 0.91	22 1.08	143 1.55	61 1.38	24 0.73
ʈ, ʈʰ, ɖ, ɖʰ		39 3.64	4 (1.22)	42 1.47	1 0.13	22 0.62	21 1.24	7 0.56
ʦ, ʦʰ, ʣ			3 (13.06)	1 (0.26)	0 (0.00)	3 0.58	1 (0.39)	0 (0.00)
ʧ, ʧʰ, ʤ, ʤʰ				16 1.00	1 0.10	48 1.07	13 0.59	29 1.64
ʈʂ, ʈʂʰ, ɖʐ					18 (12.71)	1 0.08	0 0.00	5 (1.09)
s, z						35 1.22	9 0.32	1 0.05
ʃ, ʒ							19 2.65	0 0.00
ʂ, ʐ								21 (5.86)

The co-occurrence pattern for Kalasha in Table 25 bears a strong resemblance to that of Indus Kohistani (cf. Table 22 and Table 23). First of all, combinations of two coronal obstruents that agree in both manner and retroflexion are over-attested whether they are plosives (O/E = 3.64), affricates (O/E = 12.71) or fricatives (O/E = 5.86). Moreover, most combinations that agree in both manner and non-retroflexion are also over-attested. These include combinations of two dental affricates (O/E = 13.06), two dental fricatives (O/E = 1.22) and two palatal fricatives (O/E = 2.65). Those that are not over-attested occur more-or-less as expected including pairs of dental plosives (O/E = 0.95) and pairs of palatal affricates (O/E = 1.00).

Second, combinations of two obstruents that agree in manner but disagree in retroflexion are either categorically absent or very nearly so. This includes the combination of retroflex fricatives with dental fricatives (O/E = 0.05) or palatal fricatives (O/E = 0.00); the combination of retroflex affricates with dental affricates (O/E = 0.00) or palatal affricates (O/E = 0.10); and the combination of retroflex plosives with dental plosives (O/E = 0.12). To a lesser degree, combinations of two obstruents that agree in manner but disagree in non-retroflex place features are also under-attested including the combination of dental and palatal affricates (O/E = 0.26) and the combination of dental and palatal fricatives (O/E = 0.32).

Third, plosive/sibilant combinations are relatively unconstrained with respect to place of articulation regardless of whether the sibilants are affricates or fricatives. A possible exception to this generalization is the combination of retroflex plosives with retroflex sibilants. Such combinations, which agree in retroflexion but not in manner, are under-attested whether the sibilants are affricates (O/E = 0.13) or fricatives (O/E = 0.56). This rather curious fact might be explained in one of several ways. For one thing, plosive/sibilant combinations might be subject to *retroflex dissimilation*. Under this interpretation, *Ṭ-Ç/Ç-Ṭ and *Ṭ-Ṣ/Ṣ-Ṭ combinations would be avoided in favour of those that disagree for retroflexion, such as Ṭ-Č/Č-Ṭ or T-Ç/Ç-T, and so forth. Alternatively, plosive/sibilant combinations that agree in retroflexion might be subject to *assimilation of manner*. Under this interpretation, *Ṭ-Ç/Ç-Ṭ and *Ṭ-Ṣ/Ṣ-Ṭ combinations would be avoided in favour of those that agree for manner (and retroflexion), such as Ṭ-Ṭ, Ç-Ç or Ṣ-Ṣ. In the absence of historical-comparative data to support either of these explanations, I tentatively assume that the lower-than-expected frequency of Ṭ-Ç/Ç-Ṭ and Ṭ-Ṣ/Ṣ-Ṭ combinations simply reflects the *absence of retroflex assimilation* in plosive/sibilant pairs combined with the (accidental) fact that very few Ṭ-Ç/Ç-Ṭ and Ṭ-Ṣ/Ṣ-Ṭ sequences have developed from other independent sound changes.[50]

The most significant difference between Kalasha and Indus Kohistani lies in the co-occurrence of affricates and fricatives. For ease of comparison, the figures for affricate/fricative combinations in Indus Kohistani roots from Table 23 and those in Kalasha roots from Table 25 are repeated below in Table 26. Here, as in the original tables, the relative order of consonants is collapsed. For instance, cells representing co-occurrence of dental affricates (TS) with dental fricatives (S) represent the combined figures for TS-S and S-TS sequences.

---

[50] This explanation is consistent with the account of plosive-sonorant pairs in other South Asian languages. Recall that harmonic Ṭ-Ṛ sequences are under-attested and disharmonic T-Ṛ sequences are over-attested in most Dravidian (§3.1.3) and Indo-Aryan languages (§3.2.4) (and also in most Munda languages, which are discussed in §3.4.1 below). This is assumed to reflect the absence of assimilation in these pairs, not dissimilation of place (Ṭ-Ṛ → T-Ṛ) or assimilation of manner (Ṭ-Ṛ → Ṭ-Ṭ).

Table 26 Co-occurrence of affricates and fricatives in Indus Kohistani and Kalasha roots.

Indus Kohistani (repeated from Table 23)

	S	Š	Ṣ
TS	4 (1.14)	0 (0.00)	0 (0.00)
Č	6 0.61	9 1.71	0 (0.00)
Ç̣	1 (0.25)	0 (0.00)	10 (6.23)

Kalasha (repeated from Table 25)

	S	Š	Ṣ
TS	3 0.58	1 (0.39)	0 (0.00)
Č	48 1.07	13 0.59	29 1.64
Ç̣	1 0.08	0 0.00	5 (1.09)

Both Kalasha and Indus Kohistani exhibit a general preference for coronal place agreement in affricate/fricative pairs. In each case retroflex affricates co-occur almost exclusively with retroflex fricatives and dental affricates co-occur almost exclusively with dental fricatives. Palatal affricates are exceptional; in both languages they show fewer co-occurrence restrictions than other affricates. The crucial difference between the two languages lies in the co-occurrence of palatal affricates with retroflex fricatives. While the palatal affricates of Kalasha co-occur freely with retroflex fricatives (O/E = 1.64) those of Indus Kohistani do not (O/E = 0.00). This reflects the fact that palatal affricates were subject to retroflex consonant harmony in Indus Kohistani whenever they co-occurred with a retroflex sibilant of any kind, whether affricate or fricative (cf. examples in (44)(a) above), whereas in Kalasha they were subject to retroflex harmony only when they co-occurred with a retroflex affricate but not when they co-occurred with a retroflex fricative (cf. examples in (50)(b) and (52) below). Thus Kalasha may show greater sensitivity to the relative similarity of sibilants in terms of manner than Indus Kohistani.

The lack of harmony between palatal affricates and retroflex fricatives produces a curious asymmetry in the Kalasha data that deserves mention. While palatal affricates co-occur freely with retroflex fricatives (Č-Ṣ and Ṣ-Č; O/E = 1.64), palatal fricatives do not co-occur with retroflex affricates (*Ç̣-Š and Š-Ç̣; O/E = 0.00). This suggests that palatal fricatives have served as targets of retroflex consonant harmony in affricate/fricative combinations (Š-Ç̣ → Ṣ-Ç̣) while palatal affricates have not (Č-Ṣ ↛ Ç̣-Ṣ). Stated differently, we might say that retroflex affricates have triggered harmony in fricatives but retroflex fricatives have not triggered harmony in affricates.

In sum, Kalasha exhibits a pattern of retroflex consonant harmony very similar to that of Indus Kohistani. Coronal obstruents in a root must agree for retroflexion but only if they share the same manner. However, unlike Indus Kohistani where retroflex harmony holds between sibilants of any kind, retroflex fricatives of Kalasha do not trigger harmony in palatal affricates. Thus, Kalasha exhibits a greater sensitivity to similarity within the sibilant class than Indus Kohistani. Kalasha also approaches a three-way coronal place harmony in affricate/affricate and fricative/fricative combinations, at least as a statistical tendency.

3.3.2.2 Historical-comparative evidence of consonant harmony

Historical-comparative data confirm the observations concerning retroflex consonant harmony in Kalasha. Consider the examples in (50), which demonstrate that retroflex consonant harmony has applied between obstruents of the same manner.

(50) Retroflex consonant harmony between obstruents of the same manner in Kalasha

		Kalasha		OIA	CDIAL
a.	‘dry and hard’	ɖaɖe	Pk. daɖʰa-	daːrɖʰya-, dr̩ɖʰa-	6302
b.	‘pinewood torch’	ʈʂãɖʐa	*ʧanɖʐa (ɖʐ < dr)[51]	ʧandra	4661
	‘spirit beings’	ɖʐaʈʂ	*ʤaʈʂ- (ʤ < j; ʈʂ < kʂ)[52]	jakʂa	10395
c.	‘ornate headband’	ʂuʂut(r)		*suʂuːtra-	13536
	‘nephew’	iʂpoʂi	*spaʂ-[53]	svasriːja-	13918
	‘head’	ʂiʂ		ʃiːrʂa-	12497
	‘to dry’	ʂuʂik		ʃuʂjati	12559
	‘dry, dried’	ʂuʂʈa		*ʃuʂʈa-	12555
	‘precipice’	bruʂiʂ		*bʰrãʃiʂʈʰa-	9645
	‘glue’	ʂiłeʂ	*ʃileːʂp	ʃleːʂman	12744

Roots exhibiting agreement for retroflexion between plosives are abundant in Kalasha but it is difficult to identify OIA cognates for most of them. Nevertheless, the single example in (50)(a) is consistent with the pattern found in other NIA languages where C_1 can be traced to

---

[51] The development of retroflex affricates /ʈʂ/ and /ɖʐ/ from OIA /tr/ and /dr/ appears sporadically in Kalasha but is well attested in other Dardic languages. With respect to OIA /ʧandra/ ‘shining’ Turner reconstructs Dardic *ʦanɖʐ < ʧandra. See notes under etymological group 4661 in Turner (1962–1966).

[52] /ʤ/ < /j/ is attested in the Prakrit cognate /ʤakkʰa/ < OIA /jakʂa/ ‘supernatural being’ and elsewhere in Kalasha (e.g., /ʤu/ ‘yoke’ < OIA /juga/ (CDIAL 10482) and /ʤuk/ ‘louse’ < OIA /juːkaː/ (CDIAL 10512)). /ʈʂ/ < /kʂ/ is widespread in Dardic and has already been mentioned in relation to Indus Kohistani (cf. §3.3.1.2).

[53] Cf. Turner (1962–1966) who reconstructs the progression *ʂpaʂ < *spaʂ- < *spasr- (< svasr-) for some cognates in this etymological group (CDIAL 13918). Presumably retroflex /ʂ/ in C_2 position derives from /sr/ parrallel to the development of retroflex affricates from OIA /tr/ and /dr/ in some Dardic languages.

an OIA dental plosive and C_2 to a retroflex plosive whose manner was re-enforced by means of a homorganic cluster (/-rɖʰ-/ in this case).

The examples in (50)(b) provide evidence that retroflex affricates triggered harmony in other affricates. In the case of Kalasha /ʈʂãɖʐa/ 'pinewood torch' < OIA /ʧandra/ 'shining', retroflex /ɖʐ/ developed from OIA /dr/ and subsequently triggered harmony in the preceding palatal affricate (i.e., */ʧanɖʐa/ > /ʈʂãɖʐa/). Kalasha /ɖʐaʈʂ/ 'spirit beings' derives straightforwardly from OIA /jakʂa/ 'supernatural being' by means of two other well-attested sound changes: /ʤ/ < OIA /j/ and /ʈʂ/ < OIA /kʂ/. The combination of these two developments would have produced a disharmonic sequence of two affricates (i.e., /jakʂ/ > */ʤaʈʂ/), which was then subject to retroflex consonant harmony (i.e., */ʤaʈʂ/ > /ɖʐaʈʂ/).[54]

The examples in (50)(c) are the product of retroflex consonant harmony between two fricatives. They demonstrate that both dental and palatal fricatives were potential targets and that intervening consonants did not block consonant harmony.

Recall that Kalasha exhibits an asymmetry when it comes to the co-occurrence of affricates and fricatives. This too is supported by historical-comparative data, as shown in (51) and (52).

---

[54] Cf. Morgenstierne (1973) who suggests that the retroflex /ɖʐ/ in Kalasha /ɖʐaʈʂ/ is "probably an assimilated variant of ʤ" (p. 191).

(51) Retroflex affricates trigger harmony in palatal fricatives in Kalasha

	Kalasha		OIA	CDIAL
'to learn'	ʈʂʰiʈʂ-ik	*ʃiʈʂ-	ʃikʂateː	12430
'dried fruit'	ʈʂuʈʂʰik	*ʃuʈʂ-	*ʃukʂa- (<ʃuʂka-)	12508

(52) Retroflex fricatives do not trigger harmony in palatal affricates in Kalasha

		Kalasha	OIA	CDIAL
a.	'to suck'	ʧuʂ-ik	ʧuːʂati	4898
	'to like'	ʤuʂ-ik	ʤoːʂati	5271
	'thumb, big toe'	ʤeʂʈ a˞ŋgu	ʤjeːʂʈʰa	5286
b.	'to stand'	ʧiʂʈ-ik	tiʂʈʰati	5837
	'bitter'	ʧiʂʈaka	tr̩ʂʈa	5938
	'hand span'	ʤiʂ(ʈ)	diʂʈi	6343
	'female spirit'	ʤeʂʈak	deːʂʈriː	6556

The examples in (51) demonstrate that retroflex affricates triggered harmony in non-retroflex fricatives. Interestingly, the fricatives have also become affricates in these examples but it is not clear whether this was the by-product of retroflex harmony or the product of other independent factors, such as variation between affricates and fricatives. Note that affricate/fricative sequences that agree in retroflexion are not necessarily avoided or subject to assimilation of manner in Kalasha (e.g., /ʈʂaʂa/ 'cottage cheese'; /ʈʂaʂku/ 'a kind of evergreen tree'; /ʂanɖʐu-ik/ 'to wrinkle').

The examples in (52) demonstrate that retroflex fricatives have not triggered harmony in palatal affricates regardless of whether the palatal affricate in question can be traced to OIA, as in (52)(a), or whether it results from the palatalization of OIA dental plosives, as in (52)(b). This is where the pattern of retroflex consonant harmony in Kalasha differs from that of Indus Kohistani. In Indus Kohistani, retroflex /ʂ̻/ has triggered harmony in palatal affricates whereas in Kalasha it has not. This can be seen clearly by comparing cognates such as Kalasha /ʧuʂ̻ik/ 'to suck' with Indus Kohistani /ʈ̻ʂ̻oːʂ̻áṽ/ 'to suck (out)', both of which correspond to OIA /ʧuːʂ̻ati/ 'sucks' (CDIAL 4898).

3.3.2.3 Retroflex vowels in Kalasha

A unique and intriguing aspect of the Kalasha sound system is its full set of (typologically rare) phonemic retroflex vowels. Each vowel in the system, whether oral or nasal, has contrastive retroflex and non-retroflex counterparts. The vowel phonemes of Kalasha are listed in (53).

(53) Vowel phonemes of Kalasha (Trail & Cooper, 1999; Heegård & Mørch, 2004)

FRONT		CENTRAL		BACK	
i	ĩ			u	ũ
i˞	ĩ˞			u˞	ũ˞
e	ẽ			o	õ
e˞	ẽ˞			o˞	õ˞
		a	ã		
		a˞	ã˞		

According to Heegård & Mørch (2004), the retroflex vowels are articulated with the tongue tip turned up and possibly with some bunching of the tongue body. They derive historically from the coalescence of (non-retroflex) vowels with intervocalic retroflex consonants, as shown in (54).

(54) Diachronic origins of retroflex vowels in Kalasha (Heegård & Mørch, 2004)

		Kalasha	OIA	CDIAL
a.	'kind of cheese'	kila˞	kilaːʈa-	3181
	'bent'	ko˞ho˞k	kuʈi(ka)-	3230
	'blind'	ʂe˞a	*ʃreːɖa-	12717
	'little child'	ku˞a˞k	*kuɖa-	3245
b.	'palm of hand'	pẽ˞	paːɳi-	8045
	'beads'	mã˞(h)r̃˞k	maɳi-	9731
	'arrowhead'	bõ˞	baːɳa-	9203
	'pillar'	t^hũ˞	st^huːɳaː-	13774

The examples in (54) demonstrate that intervocalic singleton retroflex consonants of OIA have been lost through lenition in Kalasha. Wherever this has happened the feature of retroflexion has been preserved and transferred to the preceding vowel. In cases where the consonant was a retroflex nasal both nasalization and retroflexion are preserved on the vowel, as shown in (54)(b).

The feature that distinguishes retroflex vowels from their non-retroflex counterparts derives historically from retroflex consonants. Thus, it is reasonable to assume that the retroflex feature on vowels and consonants is one and the same. Even so, there is no evidence that retroflex vowels participate in retroflex consonant harmony either as triggers, targets or blockers. For instance, while retroflex obstruents trigger regressive assimilation in preceding dental or palatal obstruents of the same manner class, retroflex vowels do not trigger regressive assimilation in obstruents, as shown in (55).

(55) Kalasha: Retroflex vowels do not trigger regressive assimilation in dental and palatal obstruents

		Kalasha		OIA	CDIAL
a.	‘post, column’	tʰũ˞		stʰuːɳaː-	13774
	‘dust’	udʰũ˞		*uddʰuːɖi-	2025
b.	‘nil, zero’	tsũ˞	*suːɳa-	suːna, ʃuːnja-	12567
	‘braid, plait’	ʧu˞i		ʧuːɖa-	4883
c.	‘gold’	sũ˞a		suvarɳa-	13519
	‘reed, arrow’	ʃa˞		ʃara-	12324

Moreover, Kalasha vowels do not appear to serve as targets of retroflex assimilation when they occur between two retroflex obstruents in consonant harmony domains. No phonetic study is available of the articulatory and acoustic properties of Kalasha vowels in consonant harmony domains. It would not be surprising to find that such vowels exhibit some co-articulatory retroflex colouring, which is typical of vowels in the context of retroflex consonants cross-linguistically (Hamann, 2003, pp. 111–114). However, there are no examples of *phonemic* retroflex vowels occurring in retroflex consonant harmony domains in Trail & Cooper’s (1999) dictionary. Thus, the vowels in these domains are not perceived as phonemic retroflex vowels by Kalasha speakers. This suggests that vowels do not serve as targets of assimilation even when they stand between the trigger and target of retroflex consonant harmony.

Although Kalasha vowels do not participate in retroflex consonant harmony with obstruents, they do participate in an assimilatory pattern of their own; one that could be described as retroflex vowel harmony or even vowel-consonant harmony. Consider the examples in (56).

(56) Retroflex vowel and vowel-consonant harmony in Kalasha (Heegård & Mørch, 2004)

		Kalasha		**OIA**	**CDIAL**
a.	‘to squeeze’	/pĩ˞-ik/	[pĩ˞ik] ~ [pĩ˞i˞k]	piːɖajati	8226
	‘maize bread’	/ʧa˞ha˞ka/	[ʧa˞ha˞ka] ~ [ʧa˞ha˞ka˞]		
	‘finger’	/a˞ngu/	[a˞ŋgu] ~ [a˞ŋgu˞]	aŋguli	135
	‘wind’	/sirã˞/	[sirã˞] ~ [si˞rã˞]	saraɲju	13249
b.	‘millet’	/a˞in/	[a˞in] ~ [a˞i˞ɳ]	*aɳuni, aɳu-	195, 192

The examples in (56) demonstrate that retroflexion can assimilate (optionally) from one vowel to another even across intervening consonants. Assimilation is predominantly progressive but can also be regressive, as in the case of [sirã˞] ~ [si˞rã˞] ‘wind’. In at least one case recorded by Heegård & Mørch (2004), the retroflexion of the vowel targets a dental nasal producing variation between [n] and [ɳ], as shown in (56)(b).

In summary, Kalasha exhibits two distinct patterns of retroflex assimilation, each with very different typological properties. Retroflex consonant harmony in Kalasha exhibits stringent similarity effects so that assimilation holds only between obstruents of the same manner class. All other segments, whether consonant or vowel, do not participate as triggers or targets of harmony. Moreover, assimilation between obstruents is predominantly (if not exclusively) regressive. In contrast with this, retroflex vowel or vowel-consonant harmony is primarily progressive and possibly bidirectional. Unlike consonant harmony, it does not exhibit any obvious similarity effects. Assimilation targets any vowel and can even hold between vowels and consonants. In all of these ways retroflex vowel harmony in Kalasha bears a strong resemblance to Sanskrit n-retroflexion (§3.2.1.1), which also lacks similarity effects, applies progressively and targets dental nasals.

3.3.3 The scope of retroflex consonant harmony in Dardic

The preceding sections have reviewed evidence of retroflex consonant harmony in two Indo-Aryan languages of the Dardic group: Indus Kohistani (§3.3.1) and Kalasha (§3.3.2). The full extent of retroflex consonant harmony within the Dardic group is difficult to assess. For many languages data is either unavailable or insufficient for a reliable analysis. Nevertheless, the limited data available suggests that retroflex consonant harmony of the kind observed in Indus Kohistani and Kalasha is widespread, at least as a statistical trend. Table 27 presents the co-occurrence of coronal obstruents in five Dardic languages following the convention of Pozdniakov & Segerer (2007). The results displayed here are based on counts of headwords containing word-initial $C_1V(N)C_2$ sequences in which C_1 and C_2 are coronal obstruents and N is a homorganic nasal. In order to maximize observed and expected counts for each C_1-C_2 pair the relative order of consonants is ignored and only the most essential place and manner classes are distinguished: retroflex vs. non-retroflex and plosive vs. sibilant. Once again, parentheses are used to mark values based on expected counts that are lower than 5.0.

Table 27 #$C_1V(N)C_2$ sequences in five Dardic languages[55]

Kalasha (n = 373)

	T	Ṭ	C/Č/S/Š	Ç/Ṣ
T		− −	+	
Ṭ		+ +		− −
C/Č/S/Š				− −
Ç/Ṣ				+ +

Palula (n = 113)

	T	Ṭ	C/Č/S/Š	Ç/Ṣ
T		− −	+	
Ṭ		(+ +)		(− −)
C/Č/S/Š				− −
Ç/Ṣ				(+ +)

Kalami (n = 468)

	T	Ṭ	C/Č/S/Š	Ç/Ṣ
T		− −		
Ṭ		+ +		− −
C/Č/S/Š				−
Ç/Ṣ				(+ +)

Indus Kohistani (n = 597)

	T	Ṭ	C/Č/S/Š	Ç/Ṣ
T	+	− −	+	− −
Ṭ		+ +		−
C/Č/S/Š				− −
Ç/Ṣ				+ +

Shina (n = 243)

	T	Ṭ	C/Č/S/Š	Ç/Ṣ
T		− −	+	
Ṭ		(+ +)		
C/Č/S/Š				−
Ç/Ṣ				(+ +)

For each language in Table 27, sequences of two plosives that agree in retroflexion are represented by a double "+ +" sign indicating that they are substantially over-attested. Those that disagree in retroflexion are represented by a double "− −" sign indicating that they are substantially under-attested. Pairs of plosives that agree in non-retroflexion (i.e., dental-dental

---

[55] The source of data for Indus Kohistani was Zoller (2005). Data for each of the other Dardic languages was drawn from an unpublished electronic lexical database graciously contributed by another researcher. The contributers are as follows: Ron Trail & Greg Cooper (for Kalasha; cf. Trail & Cooper, 1999); Henrik Liljegren (for Palula); Joan Baart (for Kalami); and Carla Radloff (for Shina). In order to increase counts, Radloff's Shina data was supplemented with additional data from Bailey (1924).

pairs) occur more-or-less as expected with the exception of Indus Kohistani where they too are somewhat over-attested. A similar pattern holds within the class of sibilants. Pairs of sibilants that agree in retroflexion are substantially over-attested in every case while those that disagree in retroflexion are all under-attested to varying degrees.

Plosive/sibilant pairs show a very different pattern. In most cases pairs that agree in non-retroflexion tend to be slightly over-attested (as indicated by a single "＋") while those that agree in retroflexion tend to be under-attested to some degree (as indicated by a single "−" or double "−−"). As argued earlier in relation to Kalasha (§3.3.2.1), the lower-than-expected frequency of plosive/sibilant pairs agreeing in retroflexion probably reflects the absence of diachronic retroflex harmony in these pairs rather than dissimilation of place, assimilation of manner or some other co-occurrence restriction. This is further supported by the fact that plosive/sibilant pairs disagreeing in retroflexion (i.e., T-C̣/Ṣ and Ṭ-C/Č/S/Š) are neither avoided nor preferred in most cases but occur more-or-less as expected. The only exception to this generalization is Indus Kohistani in which retroflex sibilants are under-attested in combination with plosives of any kind, whether dental or retroflex.

For each language in Table 27 it is possible to find specific examples of diachronic retroflex consonant harmony. For instance, all five languages have /s̢is̢/ 'head' corresponding to OIA /ʃiːrs̢a-/ 'head, skull'. However, Shina stands out from the other languages in exhibiting a larger number of exceptions, some of which are listed in (57).

(57) Examples of disharmonic roots in Shina

	Shina	**Indus Kohistani**	**OIA**	**CDIAL**
‘handspan’	diʈ	ɖíːʈ^{hi} ~ díʈ^{h}	diʂʈi	6343
‘to learn’	siʈʂ	ʈʂ^{h}iʈʂáṽ	ʃikʂateː	12430
‘mother-in-law’	ʃaʂ	ʂʌ̀ʂ, ʈʂòeʂ [56]	ʃvaʃruː-	12759
‘to pull’	ʐas	ʐ^{h}ʌ̀ʂ ~ ʐ^{h}ʌ̀s	??	

The examples in (57) indicate that at least some of the roots that have been subject to retroflex consonant harmony in other Dardic languages, such as Indus Kohistani, remain disharmonic in Shina. Thus, it is not clear whether the statistical trend toward agreement for retroflexion in plosive-plosive and sibilant-sibilant pairs can be attributed to consonant harmony in Shina or whether it stems from other sources such as reduplication (cf. discussion in §3.6.2).[57]

---

[56] The Dardic forms of this root probably developed from OIA /ʃvaʃruː-/ ‘husband’s or wife’s mother’ via an intermediate form such as */iʃpraʂuː/ or */ispraʂuː/, in which /ʂ/ < /ʃr/ (see CDIAL 12759 in Turner, 1962-1966). If so, Indus Kohistani /ʈʂòeʂ/ ‘mother-in-law’ may not be the product of consonant harmony since /ʈʂ/ also developed independently from /pr/ in that language (cf. Indus Kohistani /ʈʂìːʂ/ ‘flea’ but Kalasha /priʂu/ ‘flea’, both deriving from OIA */priʂu/ < Skt. /pluʂi/ (CDIAL 9029)).

[57] Carla Radloff (personal communication, May 21, 2009) told me that one of her Shina informants was inclined to say [ʐaʂ] for /ʐas/ ‘to pull’ and [ʂiʈʂ] for /siʈʂ/ ‘to learn’ but avoided doing so after it was brought to his attention. The disharmonic forms are clearly the accepted standards for these roots. Nevertheless, the anecdotal evidence suggests that some speakers may apply retroflex consonant harmony in speech and that such variants may go unreported in the literature if they are considered sub-standard. If this is so, then Shina may not differ quite so much from Indus Kohistani and other Dardic languages as the examples in (57) suggest. Zoller’s (2005) Indus Kohistani dictionary also includes a number of exceptional disharmonic forms. However, most of them are listed as having harmonic variants (see examples in (47)). Had Zoller not recorded the variation, and had the disharmonic variants been recorded as the standards, Indus Kohistani might look more like Shina. Thus, it is possible that the difference between the languages might lie more in standardization and the degree of dialectal variation reported in the literature than in the languages themselves.

In sum, agreement for retroflexion in plosive-plosive and sibilant-sibilant pairs is a widespread statistical trend among Dardic languages. Historical-comparative data confirm that retroflex consonant harmony is a contributing factor to this trend in languages such as Kalasha and Indus Kohistani. However, the same data raises doubts about the status of consonant harmony in Shina where disharmonic forms have been retained for many roots. Detailed case studies of Shina and other Dardic languages are required to determine the full extent of retroflex consonant harmony within the Dardic group.

3.3.4 A possible case of blocking in Dardic

Before moving on from the Dardic languages, it is worth pointing out what appears to be a *possible* case of blocking in Dardic retroflex consonant harmony. There is some evidence to suggest that retroflex consonant harmony might be blocked between plosives in T-ṢṬ sequences, i.e., when the retroflex plosive in C_2 position is immediately preceded by a (retroflex) sibilant. Evidence bearing on this issue is very limited because there are relatively few OIA roots/stems containing the relevant T-ṢṬ sequences. Even fewer of these can be traced to cognates in NIA languages of the Dardic group. Nevertheless, the few cognates that can be identified all show a lack of harmony whenever the intervening sibilant in a T-ṢṬ sequence is retained. Consider the examples in (58).

(58) Possible examples of blocking in Dardic T-ṢṬ sequences

		Kalasha	Palula	Kohistani	OIA	CDIAL
a.	‘to stand’	ʧiʂʈik	—	—	tiʂʈ^h^ati	5837
	‘hand span’	ʤiʂ(ʈ)	diʂʈ	ɖíːʈ^hi^, díʈ^h^	diʂʈi	6343
	‘female spirit’	ʤeʂʈak	—	—	deːʂʈriː	6556
	‘wine’	—	—	diʂʈʎʌ̃	??	??
b.	‘bitter’	ʧiʂʈaka	tríʂʈu	ʧĩʈ^hi^	tr̩ʂʈa-	5938
	‘visible; seen’	drẽʂ(ʈ)	d^h^riʂʈu	—	dr̩ʂʈa	6518
	‘written cure’	draʂʈaw	—	—	??	??

As a general rule, Kalasha and Palula have preserved OIA /ʂʈ/ and /Cr/ clusters, while Indus Kohistani (like most other NIA languages) has not. The examples in (58) suggest that, wherever /ʂʈ/ clusters have been preserved, the retroflex plosives in these clusters have not triggered harmony in any preceding dental plosives.

Kalasha has palatalized OIA dental plosives in TVṢṬ sequences whenever they occurred before front vowels (e.g., /ʧiʂʈ-/ < /tiʂʈ^h^-/). Palatalization in Kalasha entails affrication. It is possible that this sound change has bled retroflex consonant harmony. As we have seen, retroflex plosives and fricatives do not trigger harmony in palatal affricates in Kalasha. However, there is reason to believe that palatalization is not entirely responsible for the lack of retroflex harmony in these examples. First of all, palatalization has not applied to dental plosives in /Cr/ clusters (provided the /r/ has been preserved). In such cases, retroflex harmony has not applied, despite the lack of palatalization (e.g., Kalasha /drẽʂ(ʈ)/ < OIA /dr̩ʂʈa/ in (58)(b)). Secondly, retroflex harmony has not applied to TVṢṬ sequences in Palula, despite the lack of palatalization in that language. Palula is very closely related to Kalasha and exhibits a nearly identical co-occurrence pattern with respect to coronal obstruents (cf. Table 27 in §3.3.3 above). The limited data available suggests that palatalization has not occurred in Palula. Even in the absence of palatalization, retroflex consonant harmony has not applied to initial dentals in TVṢṬ sequences (e.g., Palula /diʂʈ/ < OIA /diʂʈi/ in (58)(a)). Taken together, the evidence from Kalasha and Palula in (58) suggests that retroflex harmony has not applied to plosives in TVṢṬ or TrVṢṬ sequences. The only consistent element in these sequences that might be responsible for blocking harmony is the intervening sibilant.

The Indus Kohistani cognates in (58) are also revealing. Most MIA languages and dialects simplified OIA /ʂʈ/ clusters to geminate retroflex plosives, typically with some aspiration (i.e., OIA /-ʂʈ-/ > MIA /-ʈʈ^h-/). These geminate plosives have been further simplified to singletons in most NIA languages, including Indus Kohistani. The loss of the intervening sibilant in these cases has paved the way for retroflex consonant harmony to apply between the plosives. Thus, Indus Kohistani follows the general trend: Ṭ-Ṭ^h < T-ṬṬ^h < T-ṢṬ (e.g., Kohistani /ɖíːʈ^{hi}/ ~ /díʈ^h/ 'span of hand' < MIA */diʈʈ^h-/ < OIA /diʂʈi/).[58] Significantly, in the one example where /-ʂʈ-/ is found after a coronal plosive in Zoller's (2005) Indus Kohistani dictionary, there is no retroflex harmony (e.g., /diʂʈʌ́ṽ/ 'wine' in (58)(a)).[59]

---

[58] This trend is also evident in other NIA languages with retroflex consonant harmony outside of the Dardic group. Cf. examples from Panjabi in (33)(c).

[59] The etymology of this word is uncertain. It appears to be cognate with Burushaski /diʂáo/ 'eingekochter Traubensaft [boiled grape juice]' (Berger, 1998b, p. 120). The /-ʂʈ-/ cluster may indicate that it is a loan word in Indus Kohistani.

The limited data in (58) suggests the possibility that intervening fricatives might block retroflex consonant harmony between plosives in Dardic languages. However, it is not clear that this constitutes a case of blocking in the normal sense of the word, i.e., where assimilation is unable to apply because an intervening segment obstructs the assimilatory mechanism, whatever that might be (e.g., feature spreading or feature agreement). It is possible that the apparent blocking effect in Dardic is instead a kind of similarity effect, i.e., a simple failure to trigger harmony in the first place. If long-distance assimilation is triggered only under similarity conditions, and if similarity is evaluated (at least partly) on the basis of acoustic/perceptual properties or features, then the presence of a fricative in the /-ʂʈ-/ cluster might mask or dominate the cues of the plosive to some degree, or might render the entire cluster perceptually distinct from a simple plosive. Under this hypothesis, it is not the intervening position of the fricative that matters, but only its adjacency to one of the plosives. The fricative might be expected to 'block' harmony whether it occurs before or after the plosive in C_2 position. While the Dardic languages lack consonant clusters consisting of plosive + fricative, they do have such sequences phonetically in the form of affricates. As we have seen, retroflex affricates do not trigger harmony in plosives. Thus, the languages might treat fricative-plosive clusters, such as /-ʂʈ-/, in a way comparable to affricates, such as /ʈʂ/. In both cases, the presence of an adjacent fricative element renders the plosive element sufficiently distinct from simple plosives so as to avoid the pressure for consonant harmony with other simple plosives.[60]

---

[60] This hypothesis predicts that harmony might not be blocked if the intervening fricative is non-adjacent to the surrounding plosives (i.e., T…Ṣ…Ṭ). Unfortunately, there are no roots of this type that might speak to the issue.

In sum, there is limited evidence suggesting that retroflex consonant harmony between plosives might be blocked by intervening sibilants in some Dardic languages, most notably Kalasha and Palula. However, the evidence bearing on this issue is limited and the pattern might reflect a similarity effect as opposed to a true blocking effect. The possibility of blocking is a topic that demands further research in Dardic and other South Asian languages.

3.3.5 Burushaski

Burushaski is an isolate of northern Pakistan with no established genetic affiliation. Although it is not necessarily related to the Dardic languages it is nevertheless spoken in the same general region, shares much vocabulary with them (particularly with Shina) and exhibits a similar phonemic inventory. For these reasons it is convenient to discuss Burushaski in the context of Dardic. The consonant phonemes of Burushaski are listed in (59).

(59) Consonant phonemes of Burushaski (Anderson, 1997)

LAB	DEN	RET	PAL	VEL	UVUL	GLOT
p	t	ʈ		k	q	
pʰ	tʰ	ʈʰ		k	qʰ	
b	d	ɖ		g	ɢ	
	ts	ʈʂ	ʧ			
	tsʰ	ʈʂʰ	ʧʰ			
		ɖʐ	ʤ			
(f)	s	ʂ	ʃ	(x)		h
	z					
m	n			ŋ		
	l					
	r					
w		ɻ	j			

Like most of the Dardic languages, Burushaski exhibits contrastive retroflexion across three manners of articulation within its coronal obstruent system: plosives, affricates and fricatives. It also has a retroflex approximant, IPA /ɻ/, described by Anderson as "a curious sound whose phonetic realizations vary from a retroflex, spirantized glide, to a retroflex velarized spirant" (1997, pp. 1022–1023). It derives historically from the lenition of intervocalic retroflex plosives in roots of Indo-Aryan origin, parallel to the development of retroflex flap /ɽ/ in many other languages (e.g., Bsk. /paɻo/ 'wedge' < Skt. /paːʈaka/; Bsk. /bʌɻum/ 'mare' < */vaɖam/ < Skt. /vaɖavaː/).

At least two potential forms of long-distance retroflex assimilation have been reported in the literature on Burushaski: (i) dialectal variation between harmonic and disharmonic forms of selected roots, discussed in §3.3.5.1 below, and (ii) alternations in the non-past morpheme conditioned by retroflex sibilants in verbal roots, discussed in §3.3.5.2.[61]

3.3.5.1 Root-internal dialectal variation

Retroflex consonant harmony, in the form of root-internal variation, has been reported in the Yasin dialect of Burushaski (Berger, 1974, p. 12; Anderson, 1997, p. 1040). Examples are listed in (60). Cognates from the Hunza and Nagar dialects are included for comparison.

[61] Lorimer (1935) described what might be considered a third (and highly unusual) case of long-distance retroflex assimilation in Burushaski. He noted that the retroflex approximant /ɻ/ had an effect on preceding non-adjacent velar and uvular obstruents (with the exception of voiceless velar /k/). He transcribed this with a subscript dot on velar and uvular obstruents (e.g., /g̣iɻʌs/ 'to enter'). It is not immediately obvious to me what this transcription entails in phonetic terms. At any rate, as Anderson (1997) points out, the phenomenon is either highly restricted or absent altogether today. Thus, it is not included in the present survey.

(60) Root-internal retroflex consonant harmony in the Yasin dialect of Burushaski[62]

		Yasin	Hunza/Nagar
a.	‘hole’	ɖaʈ ~ daʈ	—
b.	‘heap’	ʈʂoʈ ~ ʧoʈ	ʧʰoʈ
c.	‘briar’	ʈʂaʂ ~ ʧaʂ	ʧʰaʂ
	‘mountain’	ʈʂiʂ ~ ʧiʂ	ʧʰiʂ
	‘catch cold’	ʈʂumuʂ ~ ʧumuʂ	—
	‘ladder’	ʈʂirʈʂ ~ ʧiʂʈʂ [63]	ʧʰiʂ
	‘beam’	ʂenʈʂ	sinʈʂ

All of the examples in (60) exhibit regressive retroflex consonant harmony that is root-internal. In (60)(a) retroflex harmony holds between two plosives; in (60)(c) it holds between sibilants. Within the sibilant class, affricates and fricatives serve as both triggers and targets of assimilation. The example in (60)(b) is unique in showing harmony between a retroflex plosive and palatal affricate.

---

62 All of the Yasin examples are from Berger (1974) with the exception of /ʂenʈʂ/ ‘beam’ which is from Berger (1998b) as are all of the Hunza/Nagar examples.

63 The /r/ ~ /ʂ/ alternation in this example is independent of retroflex consonant harmony. In the Yasin dialect of Burushaski, /r/ is pronounced optionally as retroflex [ʂ] before /ʈʂ/ (e.g., /garʈʂ-/ ~ /gaʂʈʂ-/ ‘price’) and as palatal [ʃ] before voiceless palatal and velar stops (e.g., /-ɣark-, -ɣarʧa-/ [ɣaʃk, ɣaʃʧa]) (Tiffou & Pesot, 1989, p. 11; cf. Anderson, 1997, p. 1039).

An examination of data in Berger (1998b) reveals a few cases of harmony in the Nagar dialect as well. These are shown in (61).

(61) Root-internal retroflex consonant harmony in the Nagar dialect of Burushaski

		Nagar	**Hunza**	
a.	'braided hair'	ɖʐaʈi	ʤaʈi	cf. Skt. ʤaʈaː
b.	'pull'	ɖʐaʂ	ʤaʂ	cf. IK ʐʰʌʂ ~ ʐʰʌs

Here again we find an example of harmony between a retroflex plosive and palatal affricate in (61)(a), as well as an example of harmony between a retroflex fricative and palatal affricate in (61)(b).

The examples involving harmony between plosives and affricates in (60)(b) and (61)(a) are noteworthy given that examples of this kind are not found in any other case study conducted in the present survey. However, the two examples cited here are also the only examples of their kind in Berger's (1974; 1998b) Burushaski data, where disharmonic Č-Ṭ combinations are otherwise abundant. Thus, they appear to be idiosyncratic exceptions, not representatives of a general pattern in the language. In fact, the same can be said for all of the examples in (60) and (61). Disharmonic T-Ṭ and Č-Ṣ sequences are also abundant in Berger's data. Thus, the pattern of root-internal retroflex consonant harmony in (60) and (61) appears to be sporadic, not systematic, in Burushaski.[64]

---

[64] Although not presented here, counts were made of word-initial C_1VC_2 sequences in Berger's (1998b) dictionary of the Hunza and Nagar dialects of Burushaski (see Appendix B). They reveal an overall surface pattern in which pairs of obstruents agreeing in both place and manner are statistically over-attested. However, a closer examination of the data suggests that the vast majority of these pairs are the product of reduplication (e.g., /tatal/ from /tal/

3.3.5.2 Alternations in the non-past suffix [65]

The non-past suffix of Burushaski exhibits alternations that appear to be the product of harmony with a preceding retroflex sibilant in the root. According to Anderson (2007, p. 1255) the original non-past suffix may have been *-j- “or some palatalizing element”. Synchronically it can appear as a palatal sibilant such as /-ʧ-/, /-ʤ-/ or /-iʧ/ or as retroflex /-ʈʂ/. Most often it has a palatal form and also a palatalizing effect on the stem wherever the old affix has coalesced with the stem-final consonant, as shown in (62).

(62) Palatalizing effect of the non-past morpheme on final consonants (Anderson, 1997)

	Past Stem	**Non-Past Stem**
‘allow’	-sərk-	-sərʃ-
‘do’	et-	eʧ-
‘dance’	girat-	giraʃ-
‘see’	jeʦ-	jeʃ-
‘laugh’	d-aɣas-	d-aɣaʃ-
‘hide’	d-ʌɣʌj-	d-ʌɣʌʧ-

A different pattern is found after stems containing retroflex segments, as shown in (63).

‘flow slowly’; /ʈaʈaŋ/ from /ʈaŋ/ ‘cloudy’, etc.). This, combined with relatively high observed counts for some disharmonic sequences, such as T-Ṭ and Č-Ṣ, raises doubts about the extent to which retroflex consonant harmony has contributed to the surface pattern, or if it has at all.

[65] The suffix in question is described variously as “durative”, “non-past”, “imperfective” and “present” in the literature. I arbitrarily adopt the label “non-past” here.

(63) Non-past stems from roots with retroflex consonants (Lorimer, 1938; Berger, 1998b)

		Past Stem	**Non-Past Stem**
a.	'sit'	huruʈ-	huruʃ-
	'bite'	gaʈ	gaʈ-itʃ
b.	'itch'	d-ʌɣʌɻ-	d-ʌɣʌʈʂ-
	'enter'	giɻ-	giʈʂ-
	'dry'	buɻ-	buʈʂ-
c.	'eat'	ʂi- / ʂu-	ʂiʈʂ- / ʂuʈʂ-
	'come'	ɖʐu-	ɖʐuʈʂ-

Root-final retroflex plosives do not induce retroflexion in the non-past morpheme, as shown in (63)(a). Rather, the non-past morpheme triggers palatalization of the retroflex plosive, as it does for dental plosives in (62), or it surfaces as /-itʃ/ without affecting the plosive in the root.[66] In contrast with this, the non-past morpheme assimilates to a final retroflex approximant /ɻ/ in the root, as shown in (63)(b), and also to a retroflex sibilant in the root across an intervening vowel, as shown in (63)(c).

If the examples in (63)(c) are genuine cases of retroflex consonant harmony, then they run counter to the trend observed in other South Asian languages in at least two ways: (i) they involve progressive assimilation and (ii) they extend across a morpheme boundary to produce

[66] Anderson (2007, p. 1255) reports the alternation /ʈ/ → /ʈʂ/. However, the only example he cites is /huruʈ-/ → /huruʃ/ 'sit' which shows the alternation /ʈ/ → /ʃ/. This alternation is also found in Lorimer (1938) and Berger (1998b). Anderson does not discuss roots with retroflex /ʈ/ in his (1997) paper and I was not able to find examples of /ʈ/ → /ʈʂ/ in any other data source.

alternations in a suffix. Once again, however, the two examples in (63)(c) are the only examples cited in the literature. The same two examples are cited in all of the sources (Berger, 1998a, p. 131; Anderson, 2007, p. 1255; Munshi, 2006, p. 98). Thus, it is not clear whether they reflect a systematic and productive pattern in the language or whether they are idiosyncratic exceptions.[67]

In sum, the limited data available for Burushaski shows evidence of long-distance retroflex assimilation with some unique properties including: (i) assimilation between plosives and affricates; and (ii) potential cases of progressive assimilation extending across morpheme boundaries. In each case, however, examples are few and possibly sporadic in nature. Thus it is not clear at present whether they represent productive patterns or whether they reflect idiosyncratic properties of specific lexical items.

3.4 Munda

The Munda language family consists of a small group of minority and mostly non-literary languages concentrated in eastern India, predominantly in the states of Orissa, Jharkhand and Chhattisgarh. Although retroflexion is not a native feature of this family, most Munda languages now distinguish retroflex plosives and sonorants from their dental/alveolar counterparts. The consonant inventory for Mundari, shown in (64), is representative of a typical

---

[67] The examples in (63)(c) involve long-distance assimilation triggered by roots that lack a final consonant while those in (63)(b) involve local assimilation triggered by roots with a final retroflex approximant. This asymmetry suggests the possibility that the roots in (63)(c) might have had final retroflex approximants that are now lost, but whose effect is still seen in the retroflex [ʈ͡ʂ] of the past stem. If this is the case then the examples in in (63)(c) are not the product of retroflex consonant harmony but of the same local assimilation attested in (63)(b). This hypothesis is purely speculative, though it warrants further investigation.

Munda phonemic system. Aspirated consonants are typically limited to Indo-Aryan loanwords. Voiceless retroflex plosives may also be limited to Indo-Aryan loanwords in some Munda languages, but voiced retroflex /ɖ/ can always be found in native vocabulary. Retroflex sonorants occurring in Munda languages include the flap /ɽ/, the nasal /ɳ/ (in Mundari and some other North Munda languages) and the lateral /ɭ/ (only in Juang).

(64) Consonant phonemes of Mundari (Osada, 2008)

LAB	DEN	RET	PAL	VEL	GLOT
p	t	ʈ	ʧ	k	ʔ
(pʰ)	(tʰ)	(ʈʰ)	(ʧʰ)	(kʰ)	
b	d	ɖ	ʤ	g	
(bʰ)	(dʰ)	(ɖʰ)	(ʤʰ)	(gʰ)	
	s				h
m	n	ɳ	ñ	ŋ	
	r	ɽ			
		(ɽʰ)			
	l				
w			j		

3.4.1 The scope of retroflex consonant harmony in Munda

A survey of Munda languages reveals that they exhibit the same pattern of retroflex consonant harmony found in many Dravidian (§3.1.3) and Indo-Aryan languages (§3.2.4).[68] Table 28 shows the co-occurrence of coronal plosives and retroflex sonorants in four North Munda

[68] Remo may also exhibit optional nasal consonant harmony in the form of alternations. Anderson & Harrison (2008a, p. 568) report that the plural suffix /-le/ can be realized as /-ne/ if the final syllable of the stem contains a nasal consonant (e.g., /remo-le/ ~ /remo-ne/ 'people').

languages following the convention of Pozdniakov & Segerer (2007) introduced in §3.1.3. Table 29 does the same for six languages of the South Munda group. Results in both tables are based on counts of headwords containing word-initial $C_1V(N)C_2$ sequences in which C_1 and C_2 are coronal plosives or retroflex sonorants and N is a homorganic nasal.[69] Parentheses mark values based on expected counts that are lower than 5.0.

Table 28 #$C_1V(N)C_2$ 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 (n = 49)[70]

	T	Ṭ	Ṛ
T	+ +	− −	
Ṭ	− −	+ +	

Ho (Deeney, 1978; n = 178)

	T	Ṭ	Ṛ
T	+ +	−	+
Ṭ	− −	+ +	− −

Table 29 #$C_1V(N)C_2$ 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	− −	+ +	

[69] For each language in Table 28 and Table 29 the class of retroflex sonorants represented by Ṛ includes /ɽ/. In Mundari and Santali it also includes /ɽʰ/, in Kharia it includes /ɽʰ, ɳ/, and in Juang it includes /ɽʰ, ɳ, ɭ/.

[70] Data for Korwa is drawn from an unpublished electronic lexical database containing approximately 1500 words collected and transcribed by Binzy Joseph George & Christina Joseph (cf. George & Joseph, 2008).

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 Table 28 and Table 29 is essentially the same as that exhibited by Dravidian and Indo-Aryan languages with retroflex consonant harmony (cf. Table 11 and Table 20). In every case, pairs of two plosives that agree in retroflexion or non-retroflexion are substantially over-attested (i.e., Ṭ-Ṭ, T-T) while those that disagree in retroflexion are under-attested (i.e., T-Ṭ, Ṭ-T). The co-occurrence of plosives with retroflex sonorants shows just the opposite trend. Disharmonic T-Ṛ sequences occur either at or above expected frequencies while harmonic Ṭ-Ṛ sequences occur either at or below expected frequencies. Thus, all of the Munda languages surveyed exhibit a surface pattern of retroflex consonant harmony between plosives but not between plosives and sonorants.

3.4.2 Historical-comparative evidence of consonant harmony

Little is known about the history of Munda owing to a lack of literary records. Thus it is difficult to establish historical evidence of retroflex consonant harmony for these languages. The situation is further complicated by the fact that retroflex consonants occur primarily (though not exclusively) in Indo-Aryan and Dravidian loanwords or at least in pan-Indian roots of uncertain origin with cognates in Indo-Aryan, Dravidian and Munda languages. Given that many of the languages in close contact with Munda also exhibit retroflex consonant harmony it is difficult to determine whether harmony has applied independently in Munda or whether it has been imported through loanwords and multilingualism.

There is, however, comparative evidence within Munda itself to support retroflex consonant harmony. While disharmonic T-Ṭ sequences are under-attested in all of the languages surveyed, they are not under-attested to the same degree in every language. Notice that T-Ṭ sequences are only partly under-attested in Ho (Table 28) and Gorum (Table 29) (as indicated by the single "–" sign). Thus, these languages bear a resemblance to Dravidian Gondi (Table 11) and Indo-Aryan Oriya (Table 20) in which many disharmonic T-Ṭ forms are preserved alongside harmonic Ṭ-Ṭ forms. Variation between T-Ṭ and Ṭ-Ṭ is explicitly recorded for many words in Deeney's (1978) Ho dictionary. A few examples are listed in (65) along with cognates from Santali (Bodding, 1929–1936) and Mundari (Bhaduri, 1983 [1931]).

(65) Evidence of retroflex consonant harmony in North Munda languages

		Santali	Mundari	Ho
a.	'blunted arrow-head'	ʈuʈi	ʈuʈi, ʈoʈe	ʈuʈi, tuʈi
	'stump of tree'	—	ɖuʈu	ɖuʈu, duʈu
	'stick'	ɖaɳʈa	ɖaːnʈaː, ɖaːnɖaː	ɖanɖa', danɖa'
	'virgin'	—	ɖinɖaː	ɖinɖa, dinɖa
b.	'to spread the legs'	ʈanɖa	ʈanɖaː	tanɖa
	'to lean'	ʈeɳɖar	ʈenɖer	tenɖer
	'to tie in a knot'	—	ʈonɖom	tonɖom
	'tooth'	ɖaʈa	ɖaːʈaː	daʈa
	'claw of a crab'	ɖaʈɔm	ɖaɖo	daɽo

c.	‘restless, active’	taɽbəɽiə	taɽbaɽiaː	toɽoboɽi, toɽoboɽo
	‘purse’	toɽa	toɽaː	—
	‘to sting’	toɽ	tuɽ	tuː
	‘squirrel’	toɽ	tuɽa	tuː

Wherever disharmonic T-Ṭ forms are attested for Ho in (65) they correspond to harmonic Ṭ-Ṭ forms in Mundari and Santali. The examples in (65)(a) show dialectal variation between harmonic and disharmonic word forms in Ho itself and demonstrate that harmonic forms are always preferred in the other languages. Those in (65)(b) demonstrate the same preference even where 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 (65)(c) (cf. also Ho /daɽo/ in (65)(b)). The dialectal variation within Ho and the comparison of data across closely related languages suggests that retroflex consonant harmony has applied between plosives within the Munda language family.

The evidence examined here suggests that retroflex consonant harmony holds for the vast majority of Munda languages with contrastive retroflexion. No examples of long-distance retroflex assimilation extending across morpheme boundaries are reported in the literature on Munda. Thus, the pattern of retroflex consonant harmony in Munda exhibits the same typological properties found in Dravidian and Indo-Aryan languages: it is (i) root-internal; (ii) regressive (or possibly bidirectional, given the general avoidance of Ṭ-T configurations); (iii) sensitive to similarity of manner so that harmony holds only between two plosives but not between plosives and sonorants; and (iv) it does not exhibit any known blocking effects. Moreover, as in the case of Dravidian and Indo-Aryan, retroflex consonant harmony in Munda does not entail agreement for laryngeal features, as evidenced by examples such as Mundari /ɖuʈu/ 'stump of tree' (cf. Ho /ɖuʈu/ ~ /duʈu/) and Mundari /ʈenɖer/ 'to lean' (cf. Ho /tenɖer/).

3.5 Tibeto-Burman

The present survey found only a single example that *might* qualify as a case of long-distance retroflex assimilation in a Tibeto-Burman language of South Asia. Gordon & Schoettelndreyer (1970) describe a case of progressive retroflex assimilation in Sherpa, a Tibeto-Burman language of Nepal, but its status as a genuine case of consonant harmony, as defined here (§1.3.1), is doubtful. The consonant phonemes of Sherpa are listed in (66).

(66) Consonant phonemes of Sherpa (Gordon & Schoettelndreyer, 1970)[71]

LAB	DEN	RET	PAL	VEL	GLOT
p	t	ʈ		k	
pʰ	tʰ	ʈʰ		kʰ	
b	d	ɖ		g	
	ʦ		ʧ		
	ʦʰ		ʧʰ		
	ʣ ~ z		ʤ ~ ʒ		
	s		š		h
m	n			ŋ	
	l				
	lʰ				
	r ~ ʐ				
	rʰ				
w			j		

Like most Tibeto-Burman languages, retroflexion in Sherpa is contrastive only among plosives, not among sonorants. However, the single rhotic /r/ does have retroflex properties. Gordon & Schoettelndreyer describe it as an "alveolar retroflexed flap" in free variation with an "alveolar retroflexed fricative", particularly in word-initial position (1970, p. 355).

[71] The phonemic inventory of Sherpa presented here is that of Gordon & Schoettelndreyer (1970) with the exception that they treat the aspirated consonants as sequences of C + /h/. These phonetic sequences are interpreted as aspirated phonemes in other accounts such as Kelly (2004). Note that the aspirated sonorants /lʰ/ and /rʰ/ are voiceless, not breathy as they are in Indo-Aryan languages. Kelly (2004) also includes the palatal nasal /ɲ/ and a series of palatal plosives that are distinct from the so-called palatal affricates (which are phonetically palato-alveolar). Gordon & Schoettelndreyer (1970) interpret the former as /n/ + /j/ and the later as velar plosives + /j/.

According to Gordon & Schoettelndreyer (1970), Sherpa exhibits a pattern of retroflex assimilation that targets non-initial denti-alveolar consonants of any kind following word-initial retroflex plosives or /r/. Examples are listed in (67) (with tone omitted).

(67) Progressive retroflex assimilation in Sherpa (Gordon & Schoettelndreyer, 1970)

/ʈitu/	[ʈiʈu] ~ [ʈitu]	‘ask’ (interrogative)	(p. 350)
/ʈi-ni/	[ʈiɳi]	‘having asked’	(p. 354)
/ʈʰil/	[ʈʰɨɭ]	‘wrap (it)!’	(p. 356)
/rul/	[ʐuɭ]	‘snake’	(p. 356)
/ril-suŋ/	[ʐɨɭʂuŋ]	‘(she) rolled (the dough)’	(p. 366)

The examples in (67) reveal a pattern of assimilation with a number of striking properties. To begin with, assimilation is purely progressive. Secondly, it is not sensitive to the similarity of participating segments in terms of their manner of articulation. Both plosives and sonorants serve as triggers and targets. Assimilation is described as optional for plosive targets but not for sonorant targets. It may even target intervening vowels. Notice that the vowel /i/ is realized as a “high open front-centralized” variant (transcribed as [ɨ] in (67)) when it follows a retroflex consonant (Gordon & Schoettelndreyer, 1970, pp. 362-63, 366).[72] Thirdly, assimilation can extend across morpheme boundaries (e.g., /ʈi-ni/ [ʈiɳi] ‘having asked’). Finally, assimilation results in allophonic variation. That is, it produces phonetic retroflex sonorants such as [ɳ] and [ɭ], which do not occur as independent phonemes of the language.

---

[72] The vowel is transcribed as [ɩ˃] in Gordon & Schoettelndreyer (1970). Their phonetic transcriptions of [ʈiʈu] ~ [ʈitu] ‘ask’ (interrogative) and [ʈiɳi] ‘having asked’ do not include this vowel. The reason for this omission is unclear given that they define its conditioning environment as “following retroflex consonants” (p. 362).

The properties of retroflex assimilation in Sherpa set it apart from the pattern of retroflex consonant harmony observed in most other South Asian languages where assimilation is predominantly (if not exclusively) regressive, highly sensitive to the similarity of particpating segments and root-internal. The pattern in Sherpa bears a stronger resemblance to those of Sanskrit n-retroflexion (§3.2.1.1) and Kalasha retroflex vowel(-consonant) harmony (§3.3.2.3), where assimilation is predominantly progressive, lacks similarity effects and potentially extends across morpheme boundaries.

3.6 Summary and conclusion

The current chapter has presented evidence of retroflex consonant harmony in a wide range of South Asian languages. This concluding section summarizes the most important generalizations concerning the genetic and geographic scope of retroflex consonant harmony in South Asia (§3.6.1) and the typological properties associated with it (§3.6.3). In addition, some comments regarding the relation between consonant harmony and reduplication are offered in §3.6.2.

3.6.1 The scope of retroflex consonant harmony in South Asia

The present survey reveals that retroflex consonant harmony, as a static morpheme structure constraint resulting from diachronic assimilation, is widespread among South Asian languages. It applies to most Indo-Aryan languages of the Northwestern, Northern, Central and Eastern zones; most Dravidian languages of the Northern and South-Central groups; and the vast majority of Munda languages. It does not hold over Indo-Aryan languages of the Southern and Sinhalese-Maldivian zones, Dravidian languages of the South and Central groups (with the exception of Parji) or languages of the Tibeto-Burman family.

The geographic distribution of South Asian languages with and without retroflex consonant harmony follows a clear trend. Setting aside Tibeto-Burman, where there is little or no evidence of harmony, we can say that languages with retroflex consonant harmony are concentrated in the northern half of the South Asian sub-continent while those that lack retroflex harmony are concentrated in the southern half. This can be seen on the map in Figure 10, which shows the approximate location of languages with and without retroflex consonant harmony.[73] In the interest of saving space, numbers are used instead of language names on the map. A list of each language name and its corresponding number is provided in (68).

---

[73] With only two exceptions, the presence or absence of retroflex consonant harmony in each language in Figure 10 has been established through a statistical analysis of coronal co-occurrence patterns in some lexical corpus, and through historical-comparative evidence. The two exceptions are the South Dravidian languages Malayalam (②) and Kannada (③), for which no statistical analyses were conducted. The absence of retroflex consonant harmony in these languages (and in most other South Dravidian languages) is abundantly clear from historical-comparative data in Burrow & Emeneau's (1984) Dravidian etymological dictionary (DEDR). Out of 177 etymological word groups in DEDR containing items with word-initial C_1VNC_2 sequences, in which C_1 and C_2 are coronal stops or retroflex sonorants, only one Kannada word exhibits a harmonic Ṭ-Ṭ pattern: /ɖaɳɖaɳ/ 'sound of the drum called ḍavaṇe' (DEDR 2945). The retroflexion and reduplication in this item are both characterstic of onomatopoeic words in South Asian languages. All other Malayalam and Kannada words listed in the relevant etymological groups exhibit disharmonic T-Ṭ patterns or lack retroflex segments altogether (i.e., T-T). Malayalam and Kannada are included in Figure 10 to help complete the picture of south India. Together with Tamil, they are representative of the dominant 'disharmonic' pattern found in South Dravidian languages.

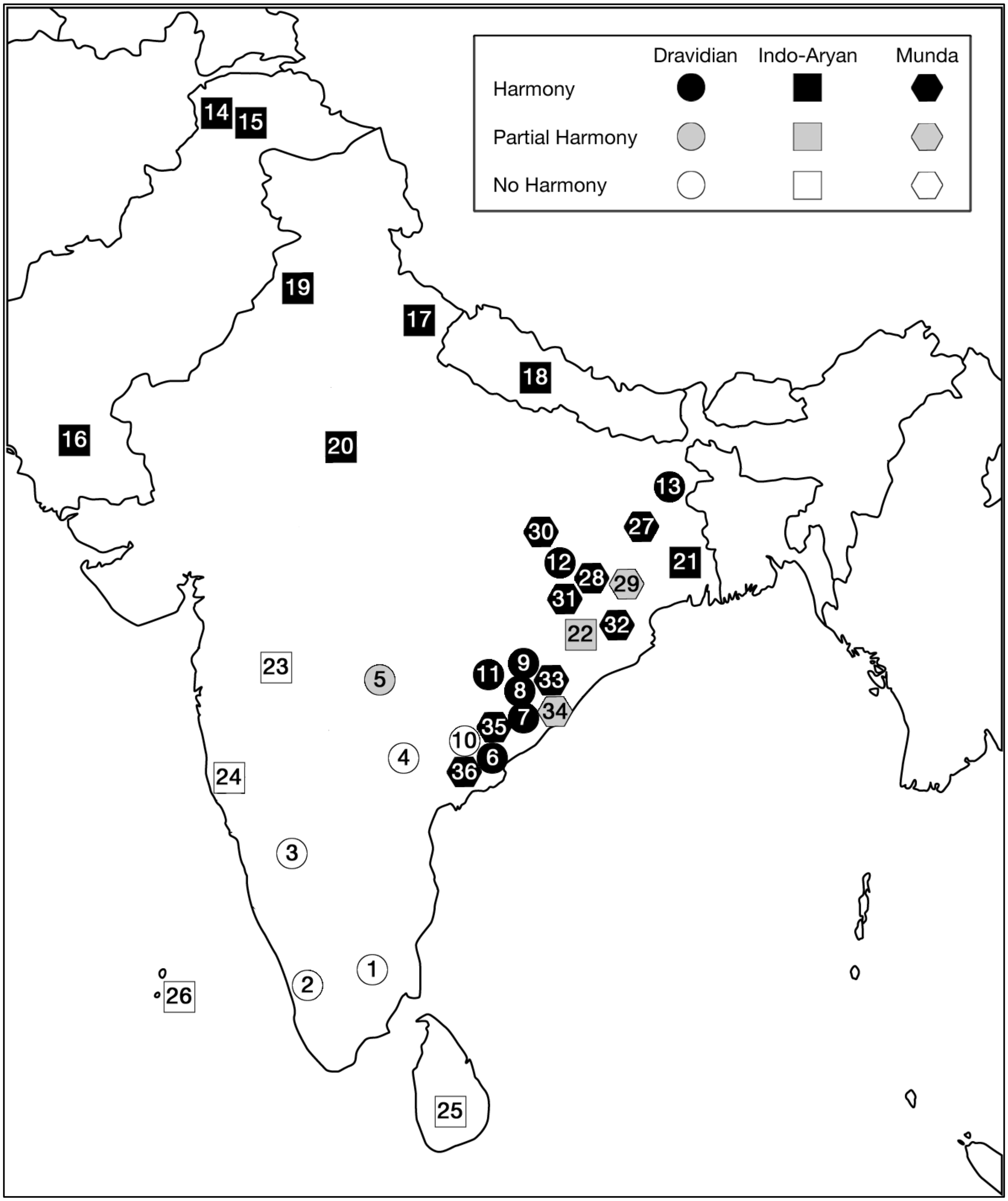


Figure 10 Approximate geographic distribution of languages with and without retroflex consonant harmony in South Asia

(68) Index of languages listed in Figure 10

#	Language	Affiliation	#	Language	Affiliation
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	I. Kohistani	IA, Northwestern	33	Remo	Munda, South
16	Sindhi	IA, Northwestern	34	Gorum	Munda, South
17	Kumauni	IA, Northern	35	Gtaʔ	Munda, South
18	Nepali	IA, Northern	36	Gutob	Munda, South

As shown in Figure 10, those Indo-Aryan languages that lack retroflex consonant harmony are precisely the southernmost ones that are in closest contact with South Dravidian languages, which also lack retroflex harmony. Those Dravidian languages that exhibit retroflex consonant harmony are precisely those in the most northerly areas, which are in close contact with Indo-Aryan and Munda languages. The Munda languages, most of which show a tendency toward retroflex consonant harmony, are concentrated in eastern parts of India where most of the Dravidian languages with retroflex harmony are also located. Thus, retroflex consonant harmony appears to be an areal property affecting most languages in the northern half of the sub-continent to some degree, except those of the Tibeto-Burman family.

3.6.2 Consonant harmony and reduplication

Morphological reduplication is a well-known typological trait of all South Asian language families (Abbi, 1992). In view of this fact it is important to note that both reduplication and consonant harmony are capable of producing the same surface pattern; one in which two non-adjacent consonants in a word agree in place and manner of articulation, as sketched in (69).

(69) Two sources of surface agreement in C_1-C_2 pairs

a. **Reduplication**	b. **Consonant harmony**
ʈa → ʈa-ʈa	taʈa → ʈaʈa

Potential distinctions between the output of reduplication and that of consonant harmony can be obscured further in cases where the stem and reduplicant are non-identical (e.g., ʈa → ʈam-ʈa or ʈu-ʈa) or where a final vowel has been elided (e.g., ʈa → ʈa-ʈa → ʈaʈ).

The pervasiveness of reduplication in South Asian languages and its ability to produce surface effects comparable to those of consonant harmony raise questions about the extent to which reduplication is responsible for the surface patterns observed in the present survey. There can be little doubt that reduplication has contributed to the surface agreement of C_1-C_2 pairs in most of the languages surveyed. However, in each case the evidence suggests that the surface pattern cannot be *reduced* to reduplication.

To begin with, words were excluded from all statistical counts if they were hyphenated in the data source to signify reduplication or some other morphological complexity. At the very least, this would have reduced the influence of reduplication throughout the study though it would not have eliminated it altogether. It is likely that not all cases of reduplication were explicitly identified by means of hyphenation in the data sources. Secondly, in the case studies of Indus Kohistani and Kalasha an attempt was made to reduce the data set to the class of unique roots in each language. To achieve this all suspected cases of reduplication were excluded whether they were explicitly identified by means of hyphenation or not. Even under these conditions the surface pattern of agreement remained the same.

Finally, and most importantly, historical and comparative data clearly support the application of retroflex consonant harmony in almost every case. Thus, while reduplication has undoubtedly contributed to the pattern of agreement observed throughout the survey it cannot be solely responsible for it. The weight of evidence indicates that retroflex consonant harmony is also a major contributing factor in all of the cases identified as such in the present study.

It remains possible that reduplication is the primary or exclusive source of surface agreement in those cases where: (i) agreement is only a tendency with numerous exceptions and (ii) historical-comparative evidence of consonant harmony is lacking. Examples of this kind might include surface agreement patterns in Berger's (1998b) Burushaski data (cf. footnote 64) and in Shina (cf. §3.3.3). It may also include agreement of non-retroflex (dental and palatal) sibilants in Dardic languages such as Indus Kohistani (§3.3.1) and Kalasha

(§3.3.2). Further research is required to determine the extent to which consonant harmony has contributed to these cases.

3.6.3 Typological properties

With few exceptions, all cases of retroflex consonant harmony reviewed in the present study exhibit a common set of co-occurring typological properties. These properties are summarized below with reference to the following parameters: the domain of assimilation (§3.6.3.1); the direction of assimilation (§3.6.3.2); the similarity of interacting segments (§3.6.3.3); and the transparency or opacity of intervening segments (§3.6.3.4).

3.6.3.1 Domain of assimilation

In South Asian languages, retroflex consonant harmony appears to be limited to the domain of the root or morpheme, where it is manifested as a static co-occurrence restriction on coronal obstruents that is the product of diachronic assimilation. Synchronic alternations do not occur except in the form of dialectal variation between conservative disharmonic root forms and innovative harmonic root forms (e.g., T-Ṭ ~ Ṭ-Ṭ). Exceptions to this trend include Sanskrit n-retroflexion (§3.2.1), retroflex assimilation in Sherpa (§3.5), and alternations in the Burushaski non-past suffix (§3.3.5.2). However, all of these exceptions exhibit other typological properties that set them apart from the dominant trend, including progressive directionality and a disregard for similarity. Moreover, it is not clear whether the Burushaski alternations are productive or sporadic.

The restriction of consonant harmony to roots is not an intrinsic property of consonant harmony systems *per se*. Rather, it is a property of retroflex consonant harmony as it occurs in South Asian languages. Elsewhere, many consonant harmony systems do produce alternations, including some that involve retroflexion, but root-internal harmony systems are very common cross-linguistically (Hansson, 2001; 2010; Rose & Walker, 2004; Rose, 2011).

3.6.3.2 Dominance and directionality

Wherever retroflex consonant harmony has applied in South Asia, roots containing two coronal obstruents of the same manner tend to agree in both retroflexion (e.g., Ṭ-Ṭ) and non-retroflexion (e.g., T-T), while those that disagree in retroflexion or non-retroflexion are avoided (e.g., Ṭ-T, T-Ṭ). As a synchronic co-occurrence restriction on roots, this pattern is ambiguous with respect to dominance (i.e., trigger-target relations) and directionality. However, historical-comparative evidence clearly reveals two trends: retroflex segments dominate other coronals and assimilation is regressive, if nothing else.

First, in all of the attested cases of diachronic assimilation, retroflex consonants are always the triggers of assimilation, never the targets, while dental consonants are always the targets, never the triggers. There is no evidence that dentals have triggered dental agreement in South Asian languages. Palatal consonants may serve as triggers of assimilation with respect to dentals (e.g., Pengo in §3.1.2, and possibly the Dardic languages in §3.3) but no clear examples of palatals targeting retroflex segments were found. In contrast to this, retroflex consonants clearly target palatal segments in some languages (e.g., Indus Kohistani in §3.3.1 and Kalasha in §3.3.2). These asymmetries can be expressed as a markedness hierarchy in which palatals dominate dentals, and retroflexes dominate all other coronals, as sketched in (70).

(70) Place of articulation hierarchy for South Asian coronals

dental 〈 palatal 〈 retroflex

The hierarchy in (70) reflects the *attested* pattern in South Asia. If the attested pattern also turns out to be the *only* pattern in the region, then the hierarchy has implications for how we interpret synchronic co-occurrence restrictions. For instance, the hierarchy implies that retroflex agreement (e.g., Ṭ-Ṭ), and possibly some cases of palatal agreement (e.g., Č-Č, Š-Š, etc.), are products of assimilation, while dental agreement (e.g., T-T) is not. Dental agreement may be nothing more than the residue of coronal consonant harmony. Roots with a T-T configuration are those that remain unaffected by retroflex and palatal assimilation.

Secondly, the vast majority of historical-comparative evidence points exclusively to *regressive* assimilation. The only examples of progressive assimilation are Sanskrit n-retroflexion (§3.2.1), Kalasha vowel(-consonant) harmony (§3.3.2.3), alternations in the Burushaski non-past suffix (§3.3.5.2), and retroflex assimilation in Sherpa (§3.5). Again, these are the same examples that exhibit other unique properties that go against the dominant trend, including alternations and/or a disregard for similarity. The study found no historical-comparative evidence of progressive assimilation in those languages where assimilation is root-internal and sensitive to similarity.

The issues of directionality and dominance are complicated by several factors. On the one hand, the regressive direction of consonant harmony might be a by-product of the fact that retroflex consonants did not occur word-initially at one time in most South Asian languages (§2.3.1). That is, we might say that harmony is purely regressive because the conditions for progressive assimilation (i.e., Ṭ-T configurations) were lacking historically. On the other hand, virtually all South Asian languages with retroflex consonant harmony have developed word-initial retroflex plosives independent of harmony.[74] As a result, Ṭ-P and Ṭ-K configurations are not uncommon. In light of this fact, the absence of Ṭ-T configurations may not be an accident of history. In theory, Ṭ-T configurations could be avoided through progressive retroflex assimilation (Ṭ-T → Ṭ-Ṭ) or regressive dental assimilation (Ṭ-T → T-T). However, as we have seen, there is no historical-comparative evidence for either of these processes (at least not in the case of root-internal, similarity-sensitive harmonies). All we can say for certain is that most South Asian languages lacked Ṭ-T configurations historically and failed to develop them alongside Ṭ-P, Ṭ-K and Ṭ-Ṭ- configurations. This may well reflect a principled avoidance of disharmonic Ṭ-T configurations, but unfortunately it does not tell us anything about dominance or the direction of assimilation.

It is worth noting one final point about directionality. Hansson (2001; 2010) has argued that regressive assimilation is a typological property of all consonant harmony systems wherever harmony is not stem controlled. Moreover, we have seen that palatal harmony in Pengo and Kuvi (§3.1.2) is strictly regressive despite the fact that the conditions for progressive assimilation are met. Thus, it is possible that retroflex consonant harmony is also strictly regressive and that it would remain strictly regressive even if Ṭ-T configurations were

---

[74] This fact is interesting in and of itself and raises other questions. For instance, does the development of word-initial retroflexes in consonant harmony domains pave the way for the development of other word-initial retroflexes? Or, does the introduction of word-initial retroflexes outside of harmony domains pave the way for retroflex consonant harmony? These questions cannot be pursued here. Suffice it to note that not all South Asian languages that admit word-initial retroflexes exhibit retroflex consonant harmony. For instance, even phonologically conservative South Dravidian languages such as Tamil permit word-initial retroflexes in onomatopoeic words and loanwords (e.g., /ʈaɳɳeɳ̪al/ 'the sound of a bell'; /ʈaːvuɳ/ < Eng. 'town'), though they do not maintain retroflex consonant harmony. Thus, it is doubtful that any correlation can be established between retroflex consonant harmony and other sources of word-initial retroflexion in South Asia.

introduced. In the end, however, this, too, is speculation. All we can say for certain is that retroflex consonant harmony in South Asia is regressive. As to whether or not it is strictly regressive, the empirical evidence is inconclusive.

3.6.3.3 Similarity effects

Retroflex consonant harmony in South Asia is highly sensitive to the similarity of participating segments in terms of their manner of articulation. With few exceptions, harmony holds only between obstruents that agree in manner, whether plosives or sibilants. Harmony does not hold systematically between plosives and sibilants or between obstruents and sonorants even when retroflexion is contrastive across all of these manner classes, as in the case of Indus Kohistani (§3.3.1). Exceptions to this generalization include Sanskrit n-retroflexion (§3.2.1), Kalasha vowel(-consonant) harmony (§3.3.2.3) and retroflex assimilation in Sherpa (§3.5). Once again, these are the same languages that exhibit one or more other unique properties, including alternations and progressive assimilation. Among those languages that exhibit root-internal regressive assimilation, most exceptions to the similarity condition are only apparent, as in the case of Dravidian (§3.1.5), or sporadic, as in the case of Burushaski (§3.3.5.1). The only systematic exception is Nepali, which exhibits retroflex agreement between plosives and sonorants. However, Nepali is also the only language in the survey in which retroflex sonorants are not phonemic. Thus, the failure to observe the distinction between retroflex plosives and retroflex sonorants may simply reflect the fact that harmony is also sensitive to phonological contrast, and that there is no phonological contrast between retroflex plosives and retroflex sonorants in the language (cf. discussion in §3.2.4).[75]

While retroflex consonant harmony in South Asia is conditioned by similarity of manner, not every manner distinction bears the same weight. Harmony is primarily sensitive to major manner class distinctions such as obstruent vs. sonorant and sibilant vs. non-sibilant. It is not necessarily sensitive to minor manner distinctions within the sibilant class. For instance, in Indus Kohistani, retroflex agreement holds between affricates and fricatives despite their different manners of articulation (e.g., Indus Kohistani /ʈ̺ʂo:ʂ̺áṽ/ 'to suck' < OIA /ʧu:ʂ̺ati/), but not between sibilants and non-sibilant plosives (§3.3.1).

In South Asia, laryngeal features play no role in conditioning retroflex consonant harmony. Agreement for retroflexion does not require or entail agreement for laryngeal features. Agreement for aspiration and breathy voice is rare in South Asia. In some Indo-Aryan languages, these features may be subject to independent dissimilatory co-occurrence restrictions (e.g., Grassman's law). Agreement for voicing under retroflex consonant harmony conditions is a statistical tendency in most cases, but remains far from categorical. Moreover, C_1-C_2 configurations that agree in retroflexion but not in voicing are typically well-attested, and

---

75 The point here is not that retroflex consonant harmony *cannot* or *should not* hold between plosives and sonorants, but only that it generally *does not* in South Asian languages, where harmony is typically conditioned by similarity. Coronal consonant harmony appears to hold between plosives and sonorants in some Australian and Nilotic languages (Hansson, 2010). However, in some Australian languages, the pattern may be more akin to retroflex consonant-vowel harmony, with retroflexion extending over a contiguous span of consonants and vowels, as opposed to a true long-distance interaction between consonants. See examples in Dixon (2002, p. 571) and Hamann (2003, p. 123). If so, then this may explain the absence of similarity effects in those cases, as local assimilation is not necessarily conditioned by similarity

examples of diachronic retroflex assimilation without agreement for voicing can be found in every case. Thus, retroflex consonant harmony in South Asia is conditioned by the similarity of interacting segments, where similarity is defined in terms of major manner classes, such as obstruent vs. sonorant and sibilant vs. non-sibilant, but not necessarily in terms of laryngeal features or minor manner distinctions within the sibilant class.

3.6.3.4 Transparency and opacity

As a general rule, intervening segments in retroflex consonant harmony domains appear to be transparent to retroflexion in South Asian languages. The study found only two exceptions to this generalization. The clearest of these is Sanskrit n-retroflexion, where retroflex assimilation is blocked by intervening coronals (§3.2.1.1). Once again, this is not entirely surprising given that Sanskrit n-retroflexion exhibits other typological properties that set it apart. Following Hansson (2001; 2010), the present study has adopted the view that Sanskrit n-retroflexion is not a true case consonant harmony as defined in much of the recent literature on the topic (§1.3.1). The other exception is found in the Dardic languages, where retroflex assimilation between non-adjacent plosives may be blocked by intervening sibilants in T-ṢṬ configurations (§3.3.4). Apart from these two exceptions, the study found no other examples of blocking in consonant harmony domains. Admittedly, most (but not all) of the case studies were based primarily on an examination of word-initial $C_1V(N)C_2$ sequences, which might not be sufficient to reveal blocking effects even if they were operative. Nevertheless, it is notable that even the vowels of Kalasha appear to be transparent to retroflex consonant harmony, despite the fact that retroflexion is contrastive on vowels in that language (§3.3.2.3).

3.6.4 Concluding remarks

In summary, retroflex consonant harmony is a widespread areal trait affecting a large number of languages in the northern half of the South Asian subcontinent, including languages from at least three of the four major South Asian families: Dravidian, Indo-Aryan and Munda. Retroflex consonant harmony in South Asia can be described in terms of three co-occurring typological properties; it is: (i) root-internal; (ii) predominantly (if not exclusively) regressive; and (iii) conditioned by similarity of manner, but not by similarity of laryngeal features. In addition, transparency of intervening segments appears to be the norm, although blocking cannot be ruled out for the Dardic languages. It is interesting to note that these typological properties tend to co-occur, and that exceptions to any one of them typically occur in systems that are exceptional in other ways. Exceptions to the dominant trend tend to exhibit their own unique set of co-occurring typological properties. The clustering of typological properties into relatively discrete co-occurring sets suggests that different assimilatory mechanisms may be at work, each with its own set of associated properties. This question is taken up in the following chapter, where it is argued that the evidence from retroflex assimilation in South Asia supports the typological distinction between long-distance consonant agreement and the serial application of local assimilation over extended domains.

Chapter 4
Mechanisms of assimilation: Feature spreading and agreement

Retroflex consonant harmony in South Asia has much to contribute to the theoretical study of consonant harmony systems. The goal of this chapter (and subsequent chapters) is to explore the implications of retroflex consonant harmony, as attested in South Asia, for phonological theories that seek to provide an account of consonant harmony systems. A central question in phonological theory concerns the mechanism or mechanisms responsible for assimilation between phonological segments. Until recently, it was widely assumed that all assimilation is the product of a single mechanism, which can be called *feature spreading* or *gesture extension*. However, recent cross-linguistic studies of consonant harmony have revealed that consonant harmony systems exhibit unique typological properties that set them apart from other patterns of assimilation. Many of these unique properties are unexpected if consonant harmony is the product of the same feature spreading mechanism responsible for other assimilation patterns. As a result, these studies have argued that most cases of consonant harmony are in fact products of a very different assimilatory mechanism; one that can be described as *feature agreement* (Rose & Walker, 2004; Hansson, 2001; 2010).

Each mechanism of assimilation, whether feature spreading or agreement, makes its own predictions with respect to at least three typological parameters: (i) the relative similarity of interacting segments; (ii) the transparency or opacity of segments that intervene between the trigger and target of assimilation; and (iii) the direction of assimilation. This chapter examines the typological distinction between feature spreading and feature agreement in light of evidence from South Asian languages. It argues that the evidence from retroflex assimilation in South Asia is largely consistent with the distinction between these two mechanisms of assimilation, and provides support for the hypothesis that consonant harmony is the product of feature agreement, not feature spreading.

The chapter draws heavily on studies by Rose & Walker (2004) and Hansson (2001; 2010). It begins by reviewing the main typological properties that distinguish consonant harmony systems from other patterns of assimilation, as identified in those studies (§4.1). The mechanisms of feature spreading and feature agreement are introduced in §4.2, along with their functional underpinnings and associated properties. Section §4.3 examines evidence from retroflex assimilation in Kalasha and other Dardic languages that supports the typological distinction between agreement and spreading, and the conclusion that retroflex consonant harmony is the product of feature agreement in these languages. A brief summary is presented in §4.4 along with some concluding remarks.

4.1 Typological properties of consonant harmony systems

Cross-linguistic studies by Hansson (2001; 2010) and Rose & Walker (2004) have identified several typological properties that are characteristic of consonant harmony systems *vis-à-vis* other patterns of assimilation. These properties have already been reviewed in some detail in §1.3.3, along with representative examples. For ease of reference, the relevant generalizations are summarized here in (1).

(1) Typological properties of consonant harmony systems (Hansson, 2001; 2010)

a. Interacting segments are constrained by similarity

b. Intervening segments are transparent to the assimilating feature

c. Assimilation shows an inherent bias toward regressive directionality

In short, consonant harmony systems are characterized by at least three typological properties: (i) the similarity of interacting segments; (ii) the transparency of intervening segments; and (iii) an inherent bias toward regressive assimilation, wherever the direction of assimilation is not stem controlled. These properties may not be absolute requirements of all consonant harmony systems, but at the very least they represent a clear typological trend. This trend stands out when consonant harmony systems are compared with other assimilation patterns, including local consonant assimilation, vowel harmony, and vowel-consonant harmony, where different properties and trends are evident. In these other systems, assimilation is typically unconstrained by similarity, blocking effects are relatively normal, and the direction of assimilation is typically determined by independent factors, such as stem control, dominance, or the distribution of perceptual cues. The following section introduces two potential mechanisms of assimilation: feature spreading (§4.2.1) and feature agreement (§4.2.2). In each case, the predictions of the mechanism are discussed, particularly as they relate to the three typological properties identified above.

4.2 Mechanisms of assimilation

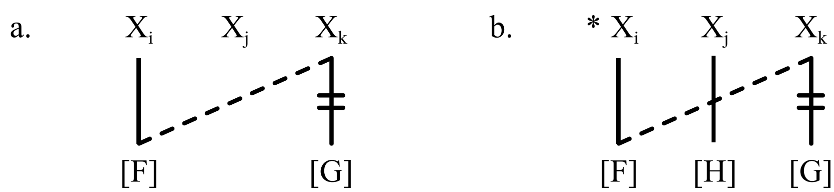
4.2.1 Feature spreading

Traditionally, phonological theories have operated on the assumption that all assimilation is the product of a single mechanism, known as *feature spreading* or *gesture extension*. According to this point of view, assimilation is accomplished by spreading a feature or extending an articulatory gesture from one segment to another. Feature spreading is regarded as a local phenomenon. That is, spreading occurs only between segments that are adjacent to one another in some respect. It is generally assumed that local feature spreading is grounded in relatively low-level articulatory factors such as co-articulation or articulatory simplification. It is often easier to maintain an articulatory feature or gesture over a span of contiguous segments than to implement a series of rapidly changing features or gestures.

Feature spreading is possible and unproblematic when the trigger and target are adjacent in the phonological string because there are no intervening segments to inhibit the spreading feature/gesture. However, cases of long-distance assimilation, such as those found in vowel and consonant harmony, are not so straightforward. In these cases, the interacting segments are not adjacent in the normal sense and the spreading feature can skip intervening segments. The apparent non-local nature of the interaction and the transparency of intervening segments present an interesting puzzle for feature spreading analyses.

Under the null hypothesis that local and long-distance assimilation are products of the same feature spreading mechanism, the apparent non-local nature of assimilation in vowel and consonant harmony has been explained by defining locality relative to some representational unit other than the segment. For instance, in autosegmental analyses, locality has been defined with respect to autosegmental tiers (e.g, Clements, 1980; Halle & Vergnaud, 1981; Poser, 1982; Steriade, 1987; Shaw, 1991). In autosegmental representations, each feature or feature class node is projected to its own independent tier and linked to segments (i.e., root nodes) by means of association lines. Feature spreading is accomplished when a feature that is linked to one segment is extended to another segment by means of a new association line. This is represented schematically in (2).

(2) Harmony as autosegmental spreading (borrowed from Hansson 2010, p. 15)



In (2), the elements X_i, X_j and X_k can represent segmental root nodes or feature class nodes such as [Coronal], [Dorsal], [Laryngeal], and so on. Similarly, the elements [F], [H] and [G] can represent feature class nodes or individual terminal features such as [−anterior], [+voice], and the like. In autosegmental representations, two features occupying the same tier are said to be adjacent, even when they are linked to non-adjacent segments (i.e., root nodes), provided that no intervening segment is associated with an element on that tier. Thus, in (2)(a), the features [F] and [G] are adjacent on their tier because the intervening segment X_j is not linked to any element on that tier. As a result, [F] and [G] can interact in a local fashion such that [F] spreads from segment X_i to segment X_k across intervening X_j (represented by the dashed association line) with concomitant delinking of [G].[1]

---

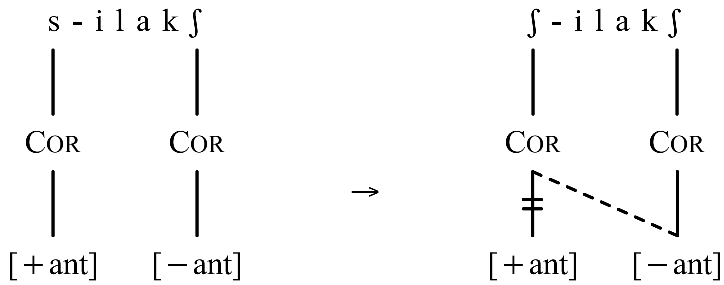
[1] Feature spreading can entail the delinking of an existing feature, as shown for [G] in (2)(a), in which case the rule is said to be *feature changing*. Alternatively, the rule can be *feature filling* if the target segment (X_k in (2)(a)) lacks the feature [G]. In this case, locality must be defined not in relation to the tier occupied by [F] and [G], but in relation to the tier that immediately dominates and hosts them. For example, spreading of a coronal feature such as [−anterior] may not require a value for [±anterior] on the target segment. In a feature filling rule it would require only that the trigger and target segments have [Coronal] nodes to host the spreading feature, and that no intervening segment has a [Coronal] node to violate adjacency on the [Coronal] tier.

The autosegmental model predicts that feature spreading can be blocked, as shown in (2)(b). In this case, the segment X_j is associated with the feature [H], which occupies the same tier as [F] and [G]. With [H] between them, [F] and [G] are no longer adjacent on their tier. The spreading of [F] from X_i to X_k across intervening X_j and [H] would violate the No-Crossing Constraint, a principle of autosegmental phonology that prohibits representations like (2)(b) in which association lines are crossed (Goldsmith, 1979). Under these conditions [F] can only target non-adjacent [G] if it first targets adjacent [H]. If [H] is resistant to assimilation, or if for some reason it can serve only as a target of assimilation and not as a trigger, then it will have the effect of blocking the spread of [F] to any potential target beyond it.

Shaw's (1991) analysis of Chumash is representative of the autosegmental approach to consonant harmony and the complications it can encounter. Chumash is a Chumashan language of California that exhibits coronal sibilant harmony between apical denti-alveolar affricates and fricatives (/ʦ, ʦʰ, ʦ', s, sʰ, s'/) and laminal postalveolar affricates and fricatives (/ʧ, ʧʰ, ʧ', ʃ, ʃʰ, ʃ'/). Segments of these two sets do not co-occur in roots or in words consisting of roots and affixes. Inputs that would violate this restriction are subject to regressive coronal assimilation. The place features of all coronal affricates and fricatives in a word are determined by those of the rightmost affricate or fricative, whether it is denti-alveolar or postalveolar.

In Shaw's analysis, denti-alveolar sibilants are distinguished by the feature [+anterior] while postalveolar sibilants are [−anterior]. Coronal assimilation is accomplished by spreading the feature [±anterior] from the rightmost affricate/fricative to all other affricates or fricatives in a word. This analysis is sketched in (3) with reference to the word [ʃ-ilakʃ] 'it is soft', in which the initial [ʃ-] derives from underlying /s-/.

(3) Coronal harmony as feature spreading under tier-based locality (based on Shaw 1991)[2]



In (3), the feature [±ant] associated with /ʃ/ spreads leftward to an adjacent [±ant] feature on the same tier, with concomitant delinking of the targeted feature. The transparency of intervening vowels and non-coronal consonants follows straightforwardly from the implications of feature geometry. These segments are not articulated with the tip/blade of the tongue. Consequently, they lack a Coronal node and coronal sub-features such as [±ant]. However, other segments present a complication for the autosegmental analysis. In addition to the coronal sibilants mentioned above, Chumash also has coronal plosives /t, t^h, t'/, nasals /n, nʔ/ and laterals /l, l'/. All of these are also fully transparent to coronal harmony; they are neither triggers, targets, nor blockers of assimilation. This is evident in the behaviour of /l/ in [ʃ-ilakʃ] 'it is soft', in (3), above.

Building on previous analyses by Poser (1982) and Steriade (1987), Shaw (1991) accounts for the transparency of non-sibilant coronals by appealing to underspecification. Although the coronal stops, nasals and laterals of Chumash could be characterized as [+ant], Shaw points out that [+ant] is redundant for these segments because they lack contrastive [−ant] counterparts. This means that they can be unspecified for the feature [±ant] in the

---

[2] The analysis sketched in (3) is that of Shaw (1991) but the example word is from Rose (2011, p. 1821).

input. As a result, they are transparent to coronal harmony at the point in the derivation where coronal assimilation applies, though they may ultimately surface as [+ant] in the output via the subsequent application of rules that supply redundant features to underspecified segments.

The autosegmental approach to harmony as feature spreading relies crucially on the notion of underspecification to account for the full range of neutral/transparent segments in languages such as Chumash. However, more recent developments in phonological theory have presented a challenge for underspecification as an explanatory device. For instance, the advent of Optimality Theory (OT) (Prince & Smolensky, 2004 [1993]) has shifted the locus of explanation from representational issues, such as underspecification, to constraint interaction. Constraint-based frameworks such as OT are output-oriented. They are concerned with constraints on output forms, not input forms. An important premise of OT is that the phonological constraints of a language should be sufficient to derive well-formed outputs regardless of the input, a principle known as Richness of the Base. According to this principle, there is no way to guarantee underspecified representations in the input. Consequently, underspecification cannot be assumed and the grammar should derive well-formed outputs even in the face of fully specified inputs.[3]

An alternative approach argues that feature spreading, or gesture extension, occurs only under strict segmental adjacency, as opposed to autosegmental tier-based adjacency. From this

[3] This challenge to underspecification arises in standard monostratal OT. However, Dresher (2009) has shown that underspecification can be accommodated within a multi-stratal model of OT. In this case, Richness of the Base applies only to the input of the first stratum, which serves as a filter to reduce representations to contrastive feature specifications. These contrastive specifications then serve as the input to the phonology proper.

point of view, the transparency of intervening segments in consonant harmony is only an illusion. A spreading feature is said to target or ‘permeate’ all segments in its path. If the spreading feature has little or no audible effect on a segment, then that segment may be perceived as unaffected by the feature/gesture and regarded as phonologically transparent to harmony (Gafos, 1999; Ní Chiosáin & Padgett, 1997; 2001).

Proponents of this view cite evidence from phonetic studies suggesting that vocalic gestures can be maintained through intervening consonants with little or no audible effect.[4] They argue that consonantal gestures are superimposed on a string of contiguous vocalic gestures. Thus, from an articulatory point of view, vocalic gestures are adjacent and uninterrupted across syllable boundaries and intervening consonants (with the possible exception of palatalized consonants or others with secondary vocalic features/gestures). This accounts for the relative frequency of vowel harmony systems cross-linguistically and the transparency of consonants in those systems.

While vocalic gestures can be contiguous across intervening consonants the inverse is not necessarily true. With few exceptions, consonantal gestures cannot be sustained across intervening vowels without significantly affecting those vowels. This is because consonantal gestures typically entail degrees of stricture that would significantly mask or obliterate any trace of the vowels. However, coronal place features/gestures constitute an important exception to this generalization. The shape and orientation of the tongue tip/blade are generally irrelevant to the production of vowels and non-coronal consonants. Thus, it is argued that coronal

---

[4] For example, see Gafos 1999, pp. 26ff and the sources cited therein.

features/gestures, including those responsible for retroflexion, are unique. Unlike most consonantal features/gestures, they can permeate intervening vowels and non-coronal consonants with little or no perceptible effect.

The analysis of Chumash coronal sibilant harmony proposed by Gafos (1999) is representative of this general approach. Gafos assumes the model of Articulatory Phonology in which the basic units of phonological representation are articulatory gestures as opposed to traditional phonological features (Browman & Goldstein, 1986; 1989; 1992). He proposes an articulatory parameter called Tongue Tip Constriction Orientation (TTCO) to represent the distinction between laminal and apical gestures. This parameter is identified as tip-down for laminal gestures (TTCO ↓) and tip-up for apical gestures (TTCO ↑). Gafos takes TTCO as the relevant parameter distinguishing Chumash sibilants on the grounds that the denti-alveolar series is described as apical and the postalveolar series is described as laminal. In the analysis proposed by Gafos, the TTCO gesture associated with the rightmost sibilant is propagated leftwards over a contiguous span of segments. This is illustrated with reference to the Chumash words /k-sunon-us/ [ksunonus] 'I obey him' and /k-sunon-ʃ/ [kʃunonʃ] 'I am obedient' in (4).

(4) Coronal harmony as gesture extension under strict locality [5]

	a.	b.
	k - s u n̺ o n̺ - u s	k - ʃ u n̻ o n̻ - ʃ
TONGUE TIP	TTCO ↑	TTCO ↓

[5] The gestural scores in (4) are adapted from Rose (2011, p. 1824) but reflect the analysis of Chumash proposed by Gafos (1999, pp. 178–184). Gafos does not present gestural scores for any Chumash words in his analysis.

In (4)(a), the apical tongue tip gesture (TTCO ↑) of the final /s/ is extended leftwards to the preceding sibilant. Likewise, the laminal tongue tip gesture (TTCO ↓) of the final /ʃ/ in (4)(b) is extended to the preceding sibilant. In both cases, the relevant gesture is extended over a span of contiguous segments. Intervening segments are not skipped. Rather, they are permeated by the gesture. Intervening vowels and non-coronal consonants are perceived as transparent because their acoustic quality is not noticeably affected by the superimposed TTCO gesture. Intervening coronal stops, nasals and liquids are necessarily affected by TTCO gestures, at least in articulation. Thus, the nasal /n/ is assumed to be articulated as apical [n̺] whenever a tip-up gesture is superimposed, as in (4)(a), and as laminal [n̻] whenever a tip-down gesture is superimposed, as in (4)(b). However, the phonetic variation of stops, nasals and liquids has no phonological significance because the language has no apical-laminal contrast for these segments. As a result, Gafos suggests that the variation goes unnoticed, thereby creating the illusion of transparency.

Proponents of the gesture extension model have argued that coronal gestures are unique. Gestures pertaining to the shape and orientation of the tongue tip/blade are possibly the only consonantal gestures capable of permeating vowels and non-coronal consonants with little or no perceptible effect. As a result, the gesture extension model predicts that coronal harmony is the only type of long-distance consonant assimilation, or at least the only type to exhibit transparency of intervening segments. Gafos (1999) regarded this as a positive prediction of the model because he was working under the faulty assumption that coronal harmony was indeed the only type of consonant harmony. However, recent studies have revealed that long-distance consonant assimilation is not limited to coronal features/gestures. Coronal consonant harmony is the most frequent type of consonant harmony, but it is certainly not the only type. Other well-attested types of consonant harmony involve the assimilation of laryngeal, nasal, liquid and dorsal features/gestures (Hansson, 2001; 2010; Rose & Walker, 2004; Rose, 2011).

The transparency of intervening segments is not limited to cases of coronal consonant harmony. Many other consonant harmony systems exhibit transparency effects in which intervening segments are clearly not targeted or permeated by the harmonizing feature/gesture. For instance, recall that the Bantu language Yaka exhibits nasal consonant harmony in which the perfective suffix surfaces as [-ini] or [-ene] in stems containing nasals, and as [-idi] or [-ele] elsewhere (§1.3.3.2). Intervening consonants and vowels are clearly unaffected by nasalization. Thus, the stem /-mí:tuk-ili/ 'sulk' is realized as [-mí:tuk-ini], with nasalization of the consonant in the suffix, but not *[-mĩ́:nũŋ-ĩni], with nasalization of all intervening vowels and consonants, as we might expect under the gesture extension model (Hansson, 2010, p. 86).

Similarly, in Tlachichilco Tepehua, dorsal consonant harmony operates over non-adjacent velar-uvular sequences to produce uvular-uvular sequences (/k…q/ → [q…q]). This language also has a pattern of local assimilation in which the high vowels /i, u/ are lowered to [e, o] when they are adjacent to uvular consonants. However, these vowels are not lowered when they occur between uvular consonants in harmony domains, provided they are not immediately adjacent to one of those uvular consonants (i.e., /k…i…q/ → [q…i…q], not *[q…e…q]) (Hansson, 2010, pp. 163–164). This is unexpected if the dorsal feature/gesture responsible for uvularization and vowel lowering is extended over all segments in the harmony domain. Thus, it is not clear that gesture extension can account for the full range of attested consonant harmony systems or the transparency of segments in those systems.

Another important criticism that has been levelled against spreading models is their general failure to predict similarity effects (Hansson, 2001; 2010; Rose & Walker, 2004). For instance, recall that Chumash coronal harmony holds only between coronal sibilants while coronal plosives, nasals and liquids are neither triggers, targets, nor blockers. In autosegmental spreading accounts, such as Shaw (1991), this is attributed to the fact that the spreading feature (i.e., [±ant]) is only contrastive for coronal sibilants and, hence, only specified on segments of that class. The class of participating segments is that which is contrastively specified for the spreading feature. The fact that sibilants are more similar to each other than to plosives or sonorants is irrelevant. It follows from this that we should expect coronal stops, nasals and liquids to interact with the sibilants if they too had contrastive [+ant] and [−ant] counterparts. However, this prediction is not always borne out. In §4.3.1.2 below, I review the evidence from the Dardic languages of South Asia presented in Chapter 3. In those languages, retroflex assimilation applies between two plosives or two sibilants, but not between plosives and sibilants, despite the fact that both segment types have contrastive retroflex and non-retroflex counterparts. In such cases, the class of participating segments cannot be defined as the set of segments that is contrastive for retroflexion. Rather, assimilation applies only between segments that are contrastive for retroflexion *and* highly similar in terms of other features. Although it might be possible to derive effects like this in a spreading analysis by stipulating feature restrictions on the class of triggers and targets, there is nothing in the spreading mechanism *per se* to predict or demand similarity effects. This point is significant in light of the fact that similarity effects are the norm in consonant harmony systems, not the exception. If all assimilation is the product of feature spreading then similarity effects are not expected in consonant harmony any more than they are in local consonant assimilation.

Feature spreading is not associated with any particular directional bias. This is appropriate for patterns of local assimilation, and possibly for many cases of vowel and vowel-consonant harmony, where the direction of assimilation is determined by independent factors such as stem control, dominance, and the distribution of perceptual cues. However, spreading models do not predict or explain the bias toward regressive assimilation in consonant harmony systems.

In summary, feature spreading models have adopted the null hypothesis that all assimilation is the product of a single mechanism, regardless of whether it applies to segments that are adjacent or non-adjacent in the phonological string. These models provide a natural account of blocking effects and other properties commonly associated with local assimilation patterns. However, under feature spreading accounts, the unique typological properties associated with consonant harmony systems are either problematic (as in the transparency of intervening segments) or at least unexpected (as in the case of similarity effects and the bias toward regressive assimilation). Thus, while feature spreading provides a natural account of local assimilation, and possibly of many vowel and vowel-consonant harmony phenomena, it does not predict the particular set of properties that are most characteristic of consonant harmony systems.[6]

[6] For a more recent attempt to provide a unified account of local and long-distance assimilation within a feature spreading framework, see Jurgec (2010). Jurgec's dissertation became available at a rather late stage in the development of the present study. For this reason, it is not addressed here.

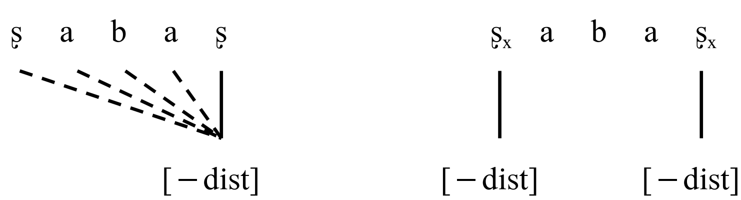
4.2.2 Feature agreement

While spreading models are generally successful in accounting for local assimilation, they fall short in accounting for some of the most characteristic typological properties of consonant harmony systems. In response to this, recent studies have argued that feature spreading is not responsible for all cases of assimilation. These studies argue that at least some cases of long-distance assimilation are the product of a different mechanism; one that can be described as *feature agreement*. This mechanism has been modelled using output-output correspondence within Optimality Theory, in what has come to be known as the Agreement By Correspondence (ABC) model of long-distance consonant agreement (Rose & Walker, 2004; Hansson, 2001; 2010). Formal details of this model are discussed in Chapter 5. For the present it is sufficient to sketch the general concept of feature agreement along with its functional grounding and associated typological properties.

The distinction between long-distance feature agreement and local feature spreading is represented schematically in (5), with reference to the Kalasha word /ʂabaʂ/ 'congratulations!'.

(5) Local feature spreading vs. non-local feature agreement

a. Spreading b. Agreement



Whereas local feature spreading involves the extension of a feature or gesture over a contiguous span of segments, as shown in (5)(a), feature agreement involves the repetition of a feature or gesture on non-contiguous segments, as shown in (5)(b). Feature agreement models operate on the premise that phonological similarity can form the basis for non-local interaction between segments in an output string. That is, a formal relation can be established between non-adjacent segments on the basis of their similarity to one another. This is indicated in (5)(b) by means of co-indexation. Segments that participate in such relations can be subject to assimilation (and possibly other phonological constraints) without regard to intervening segments. Thus, intervening segments in (5)(b) are truly transparent. They are completely ignored and skipped by the agreeing feature, not permeated by it as in (5)(a).

Proponents of this approach argue that similarity-based agreement is grounded in the psycholinguistic domain of speech planning. In support of this hypothesis, they cite parallels between consonant harmony systems and speech errors. In speech production, near-identical sounds often interfere with and intrude upon each other. This can be seen, for instance, in the mispronunciation of the English phrase *subjects show* as *shubjects show*, where the intended sequence [s…ʃ] is realized as [ʃ…ʃ]. Such 'slips of the tongue' are common cross-linguistically and exhibit many typological affinities with consonant harmony systems. Like consonant harmony, speech errors of this type exhibit the following properties: (i) they are highly sensitive to the similarity of interacting segments; (ii) the interacting segments are non-adjacent and intervening segments are ignored (i.e., transparent); and (iii) assimilation is predominantly anticipatory or regressive with respect to direction (Rose & Walker, 2004, pp. 487–490, Hansson, 2010, chapter 6).[7] The parallels between consonant harmony and speech errors

---

[7] Hansson (2010) identifies a 'palatal bias' as another property common to both consonant harmony systems and speech errors. This term refers to the trend in which posterior coronals (e.g., 'palatals' such as /ʧ/ and /ʃ/) are most often the triggers of assimilation/intrusion while anterior coronals (e.g., denti-alveolars such as /t/ and /s/) are most often the targets. However, the same palatal bias is also evident in many patterns of local assimilation. Thus, it is

suggest that both arise out of the same functional grounding. In effect, consonant harmony can be regarded as the phonologization of the mechanism responsible for similarity-induced speech errors.

Similarity-induced interactions of any kind can been explained in terms of spreading-activation models of language production. In these models, the activation of a phonological feature or gesture entails the activation of associated processing nodes. When two consonants in a word are highly similar, there is significant overlap in the nodes that are activated. This creates the potential for interference between similar segments in language production (see Rose & Walker, 2004, pp. 488–499; Hansson, 2010, p. 340; and the sources cited therein).

The trend toward anticipatory interference has been attributed to the functional requirements of serial-order production. Three functional requirements are recognized: (i) a 'turn-on' function, in which the system must activate the present; (ii) a 'turn-off' function, in which the system must deactivate the past; and (iii) a 'prime' function, in which the system must prepare to activate the future. The 'prime' and 'turn-on' functions are largely concurrent. As present elements are being activated and implemented, future elements are being planned. Thus, the planning of future activations can easily interfere with present activations, resulting in anticipatory intrusions (Dell, Burger, & Svec, 1997; Hansson, 2010).

---

not something that distinguishes consonant harmony and speech errors from other patterns of assimilation. For this reason I omit it from the present discussion.

In summary, recent studies have argued that all cases of assimilation are not necessarily products of the same mechanism of assimilation. While some cases are the product of feature spreading others are the product of feature agreement. Unlike feature spreading, feature agreement is contingent upon the phonological similarity of participating segments and intervening segments are genuinely transparent. If it is grounded in the psycholinguistic domain of speech planning, as argued by Hansson (2001; 2010), then there is also good reason to expect feature agreement to show an anticipatory or regressive bias with respect to the direction of assimilation. Thus, feature agreement is associated with those typological properties that are most characteristic of consonant harmony systems, including: (i) similarity effects; (ii) transparency effects; and (iii) a bias toward regressive directionality.

4.2.3 Summary of mechanisms and associated properties

The preceding sections reviewed two mechanisms of assimilation that have been proposed in the phonological literature: feature spreading and feature agreement. Each mechanism can be associated with a different set of typological properties. Feature spreading is expected to operate locally (i.e., between adjacent segments) while feature agreement is expected to operate long-distance (i.e., between non-adjacent segments), at least potentially. Apart from this, the remaining typological properties associated with each mechanism are summarized in Table 30.

Table 30 Typological properties associated with long-distance feature agreement and local feature spreading

	Agreement	Spreading
Similarity	Interacting segments share a high degree of similarity	Interacting segments are those that can accommodate the spreading feature
Opacity	Intervening segments are transparent non-undergoers	Intervening segments are undergoers, opaque blockers, or both
Direction	Assimilation is primarily regressive	No inherent/default directional bias

Feature spreading and feature agreement have different functional motivations. Local feature spreading is often attributed to low-level articulatory factors such as co-articulation or articulatory simplification. As a result, it is often insensitive to cognitive factors such as phonological contrast and similarity. Feature agreement, however, is grounded in higher-level cognitive factors. More specifically, agreement stems from the interference that can arise in producing one segment while at the same time anticipating the production of another highly similar segment. As a result, it tends to exhibit an anticipatory bias and a sensitivity to cognitive factors such as phonological contrast and similarity.

4.3 Evidence from Kalasha and other Dardic languages

The typological distinction between feature agreement and feature spreading is largely consistent with the evidence from retroflex assimilation in South Asian languages reviewed in Chapter 3. In the vast majority of cases, where retroflex assimilation affects segments that are non-adjacent in the phonological string, the pattern of assimilation exhibits properties consistent with feature agreement: assimilation is regressive, interacting segments are conditioned by similarity, and intervening segments do not block assimilation. Some exceptions can be found. However, as noted in §3.6.4, the typological properties of retroflex assimilation patterns in South Asia tend to cluster into two relatively discrete sets. In most cases, if a pattern violates the dominant trend with respect to one parameter, it also violates the trend with respect to other parameters. For example, the few cases of retroflex assimilation that are progressive also appear to be unconstrained by similarity. This is true of n-retroflexion in Sanskrit (§3.2.1.1), retroflex vowel-consonant harmony in Kalasha (§3.3.2.3), and progressive retroflex assimilation in Sherpa (§3.5). Thus, the properties associated with each mechanism of assimilation tend to cluster together. This is consistent with the hypothesis that two independent mechanisms of assimilation are at work, each with its own functional grounding and associated typological properties.

This section focuses on evidence from Kalasha, an Indo-Aryan language of the Dardic group that provides compelling support for the typological distinction between feature spreading and feature agreement. Kalasha is particularly relevant to this issue because it has a rich inventory of retroflex phonemes and because it exhibits at least three independent patterns of retroflex assimilation. One of these patterns has typological properties consistent with feature agreement while the others have properties consistent with feature spreading. The co-existence of both types of assimilation in one language provides support for the hypothesis that two different mechanisms of assimilation are at work, and that consonant harmony is the product of agreement, not spreading.

Three patterns of retroflex assimilation in Kalasha are discussed below: retroflex consonant harmony (§4.3.1); local retroflex consonant assimilation (§4.3.2); and retroflex vowel-consonant harmony (§4.3.3). The first and last of these have been described already in §3.3.2, as part of the survey of retroflex harmony systems in South Asia. Relevant details and examples are repeated here for ease of reference, along with some details concerning cognate patterns in Indus Kohistani and Palula, which serve to clarify and highlight the properties of Kalasha.

4.3.1 Retroflex consonant harmony

Kalasha and other Dardic languages exhibit a pattern of retroflex consonant harmony involving coronal obstruents. This section examines the typological properties of this pattern with respect to the direction of assimilation (§4.3.1.1), the similarity of interacting segments (§4.3.1.2), and the transparency and/or opacity of intervening segments (§4.3.1.3).

4.3.1.1 Regressive directionality

Retroflex consonant harmony in Kalasha and other Dardic languages is realized as a static morpheme structure constraint, not as a productive morpho-phonological alternation. As a result, it is difficult to determine directionality from a synchronic point of view. However, historical-comparative data clearly indicates that retroflex consonant harmony has applied regressively in Kalasha and in all other South Asian languages with retroflex consonant harmony reviewed in Chapter 3. In every case, the diachronic development of retroflex consonant harmony follows the pattern T-Ṭ > Ṭ-Ṭ. A few representative examples from Kalasha are repeated in (6) below (cf. §3.3.2.2).

(6) Kalasha: Historical evidence of regressive consonant harmony (T-Ṭ > Ṭ-Ṭ)

		Kalasha		**OIA**
a.	‘dry and hard’	ɖaɖe	< daɖʰa-	< daːrɖʰja-, dr̩ɖʰa-
b.	‘spirit beings’	ɖʐaʈʂ	< *ʤaʈʂ-	< jakʂa-
c.	‘head’	ʂiʂ	< *ʃiʂ	< ʃiːrʂa-

All cases of retroflex consonant harmony in South Asia include examples of the type T-Ṭ > Ṭ-Ṭ, in which a non-retroflex coronal assimilates to a following retroflex consonant. This pattern may result from one of two asymmetries: (i) a directional asymmetry, in which assimilation is strictly regressive; or (ii) a trigger-target asymmetry in which retroflex segments always dominate non-retroflex coronals. In principle, the question of which asymmetry is responsible for the pattern could be resolved by examining the treatment of input sequences of the type Ṭ-T. If assimilation were strictly regressive and non-retroflex coronals were potential triggers, then input Ṭ-T would map to output T-T, with regressive de-retroflexion. Alternatively, if progressive assimilation were possible and retroflex consonants were always dominant, then input Ṭ-T would map to output Ṭ-Ṭ, with progressive retroflexion.

Unfortunately, the evidence required to resolve this issue is lacking because, as a general rule, sequences of the type Ṭ-T did not occur for independent historical reasons in most South Asian languages (§2.3.1), and languages such as Kalasha have failed to develop them despite the fact that they have developed Ṭ-P and Ṭ-K configurations. This may reflect a principled avoidance of Ṭ-T configurations, but it does not tell us anything about the direction of assimilation. It is worth noting, however, that the data is consistent with strictly regressive assimilation, which is independently attested in palatal consonant harmony systems in Pengo and Kuvi (§3.1.2.2). In these languages, input T-Č maps to output Č-Č (with regressive palatalization) but input Č-T is preserved faithfully without progressive palatalization (Č-T ↛ Č-Č) or regressive de-palatalization (Č-T ↛ T-T).

In sum, historical-comparative evidence indicates that retroflex consonant harmony in Kalasha and other South Asian languages is clearly regressive. It remains unclear whether it is *strictly* regressive or whether it also operates progressively. However, the data available at present is consistent with strict regressive directionality. The viability of this interpretation is supported by evidence from other coronal consonant harmony systems in the region, most notably from palatal consonant harmony in Pengo and Kuvi.

4.3.1.2 Similarity effects

Kalasha and other languages of the Dardic group provide some of the most striking examples of similarity effects involving retroflex consonant harmony. Outside of the Dardic group, all previously identified cases of retroflex consonant harmony can be classified as exhibiting assimilation between sibilant or non-sibilant coronals, as shown in Table 31.

Table 31 Cases of retroflex consonant harmony reported in Rose & Walker (2004) and Hansson (2001; 2010) classified as sibilant or non-sibilant harmony

	Language (Genetic Affiliation)	Retroflex Consonant Harmony	
		Non-Sibilant	Sibilant
a.	Malto (Dravidian)	✓	–
	Javanese (Austronesian)	✓	–
	Pohnpeian (Austronesian)	✓	–
	Gaagudju (Australian)	✓	–
	Gooniyandi (Australian)	✓	–
	Mayali (Australian)	✓	–
	Murrinh-patha (Australian)	✓	–
b.	Benchnon Gimira (Omotic, Afro-Asiatic)	–	✓
	Capanahua (Panoan)	–	✓
	Kinyarwanda (Bantu)	–	✓
	Komi-Permyak (Finno-Ugric)	–	✓
	Nebaj Ixil (Mayan)	–	✓
	Rumsen (Costanoan, Penutian)	–	✓
	Wanka Quechua (Quechuan)	–	✓

Table 31 lists all cases of retroflex consonant harmony (outside of the Dardic group) reported in recent surveys by Rose & Walker (2004) and Hansson (2001; 2010).[8] All of the languages in the upper half of the table, identified as group (a), exhibit retroflex consonant harmony between non-sibilant coronals. In most cases the harmonizing segments are plosives, although in the Australian languages they also include sonorants. All of the languages in the lower half of the table, identified as group (b), exhibit retroflex consonant harmony between coronal sibilants (i.e., affricates and fricatives). No language in Table 31 exhibits retroflex

[8] Hansson's (2010) survey includes the Dardic language, Kalasha, citing preliminary results from the present study that were reported in Arsenault & Kochetov (2009; 2011). Since my intent here is to compare evidence from the present study against that of previous studies, I have omitted Kalasha from Table 31.

consonant harmony between sibilant and non-sibilant coronals. This might be interpreted as a similarity effect in which retroflex consonant harmony is sensitive to the sibilant vs. non-sibilant distinction. However, data from the languages in Table 31 is ambiguous on this point because all of the languages in group (a) lack retroflex sibilants while those in group (b) lack retroflex non-sibilants. Thus, with few exceptions, the class of segments participating in retroflex harmony is coextensive with the class of segments that is contrastive for retroflexion in each language. Under these conditions it is impossible to say whether retroflex consonant harmony is conditioned by similarity or whether it simply operates over all segments that are contrastive for the harmonic feature.[9]

Unlike the languages in Table 31, Kalasha and other Indo-Aryan languages of the Dardic group provide unambiguous evidence of similarity effects respecting the sibilant vs. non-sibilant distinction in retroflex consonant harmony. These languages maintain contrastive retroflexion in both sibilant and non-sibilant obstruents. Retroflex plosives, affricates and fricatives are all distinguished from dental counterparts (i.e., /ʈ, ʈʂ, ʂ/ vs. /t, ʦ, s/, with further laryngeal distinctions), while affricates and fricatives are also distinguished from palatal counterparts (i.e., /ʧ, ʃ/, etc.). The Dardic languages constitute the only known examples of

---

[9] Komi-Permyak and Malto are exceptions to this generalization. Strictly speaking, it is innaccurate to say that the class of segments participating in retroflex harmony is coextensive with the class of segments that is contrastive for retroflexion in these languages. Komi-Permyak appears to exhibit similarity effects *within* the sibilant class (i.e., sensitivity to the affricate vs. fricative distinction; Kochetov, 2007; cf. Hansson, 2010, p. 54). Malto may exhibit similarity effects *within* the non-sibilant class (i.e., sensitivity to the sonorant vs. obstruent distinction; see §3.1.1), if the flap /ɽ/ can be considered the retroflex counterpart of /r/. However, like all of the other languages in Table 31, Komi-Permyak and Malto do not exhibit contrastive retroflexion in both sibilant and non-sibilant classes. Thus, they do not provide unambigous evidence for similarity effects respecting that particular distinction, which is the focus of the current discussion.

retroflex consonant harmony in phonological systems of this kind (i.e., in systems with contrastive retroflexion in both sibilant and non-sibilant obstruents). These systems provide unambiguous evidence of similarity effects in retroflex consonant harmony. In the Dardic languages, retroflex consonant harmony holds only between obstruents of the same manner class. For example, in Indus Kohistani (§3.3.1), harmony operates over plosive pairs (7) and sibilant pairs (8), but not over mixed plosive-sibilant or sibilant-plosive pairs (9).

(7) Indus Kohistani: Retroflex consonant harmony between plosives

a.	toːtáː	'butterfly'	dʰʌtrìː	'burnt food'
	tʌ̀tʰ	'hot; heat'	tùnd̥	'a kind of basket'
	tʰatʌ́r	'smallpox'	dùnd̥	'a flock, herd'
	dítʰⁱ	'given'	dʰʌndàʰ	'dealings, business'
b.	ʈʌʈúː	'a small horse'	ɖʰàːɖ̊	'a woodpecker'
	ʈàːʈʰ	'a small rug'	ʈaɳɖáṽ	'to beat'
	ʈʰaʈʌ́r	'shallow'	ɖáːɳɖ̊	'a stick'
	ɖíːʈʰⁱ	'span of hand'	ɖʰã́ːɳɖ̊	'a pond'

c. *t…ʈ, *ʈ…t, *t…ɖ, *ɖ….t, *d…ɖ, *ɖ…d, etc. (no retroflexes with non-retroflexes)

(8) Indus Kohistani: Retroflex consonant harmony between sibilants (affricate and fricative)

a.	tsítsʰⁱ	'nipple, breast'	tsàs	'a pinch'
	tsã̀z	'soft'	sʌzúː	'sister's son'
	zʰã́ːz	'a branch of a holm oak'	zʰʌnzéːr	'a kind of bird'

b.	tʃiːtʃʌ̀kʰ	'smallpox'	tʃãːtʃúː	'a dwarf'
	tʃuʃtìː	'absorption'	ʃã̀ːtʃ	'a kind of mungo'
	tʃʰʌ̃ʒⁱ	'a winnowing tray'	ʃiʃáṽ	'to dry (up)'
	ʃòːʃaː	'decoration'	ʃʌmʃʌtáː	'a turtle'
c.	ʈʂìʈʂʰ	'grey, spotted'	ʈʂʰiʈʂáṽ	'to learn'
	ʈʂoːʂáṽ	'to wring out'	ʈʂʰʌnʐòː	'a curry comb'
	ʐàːʈʂ	'a grape'	ʐʌmʈʂúː	'a son-in-law'
	ʂìʂ	'a head'	ʂùːʂ	'decent, fine, proper'

d. *ts…ʈʂ, *ʈʂ…ts, *tʃ…ʈʂ, *ʈʂ…tʃ, *ts…ʂ, *ʂ…ts, *tʃ…ʂ, *ʂ…tʃ, *s…ʂ, *ʃ…ʂ, *ʂ…ʃ, etc.
(no retroflexes with non-retroflexes)

(9) Indus Kohistani: No retroflex consonant harmony between plosives and sibilants

a.	tsaʈáṽ	'to lick'	ʈèːtsʰ	'a flint'
	siʈìː	'a whistle'	ʈʰosàː	'a fist, punch'
	taʈʂʰáṽ	'to carve'	ʈʂʰatáṽ	'to plaster'
	dùːʂ	'a sin'	sáːɳɖ̊	'a bull'
b.	tʃʌʈúː	'a grater for spices'	tʃíːɳɖ̊	'a crack, fissure'
	ʃòʈʰ	'a bump, swelling'	ʃʌ̀ɳɖ̊	'barren, castrated'

In the case of Indus Kohistani, the set of segments that interact with each other in consonant harmony is not coextensive with the set of segments that is contrastive for retroflexion. Retroflexion is contrastive in all obstruent classes, whether sibilant or non-sibilant. However, only those obstruents that agree in manner along the sibilant/non-sibilant dimension interact with each other when it comes to retroflex harmony. Obstruent pairs that disagree in manner are not subject to harmony.

Kalasha (§3.3.2) exhibits the same overall pattern as that of Indus Kohistani but differs in the treatment of certain sibilant pairs. As in the case of Indus Kohistani, retroflex consonant harmony in Kalasha operates over plosive pairs, as shown in (10), and sibilant pairs, whether both sibilants are affricates, as in (11), or fricatives, as in (12). Harmony does not apply between sibilants and non-sibilants, as shown in (13).

(10) Kalasha: Retroflex consonant harmony between plosives

a.	da˞u tatu	'festival of beans'	dodak hik	'to wait'
	tʰedi	'now'	dʰenta	'mountain (of rock)'
	dit	'half full'	dond	'bull'
b.	ʈoʈ	'apron'	ɖinʈa	'efficient'
	ʈʰeʈ karik	'to scatter'	ɖonɖ	'double bride-price'
	ɖuɖ-ik	'to sleep'	ɖunɖulaʈ	'village of Dundulet'

c. *t...ʈ, *ʈ...t, *t...ɖ, *ɖ....t, *d...ɖ, *ɖ...d, etc. (no retroflexes with non-retroflexes)

(11) Kalasha: Retroflex consonant harmony between sibilant affricates

a.	ʦẽʦaw	'squirrel'	ʦurʦun-ik	'to become weak'
b.	ʧʰaʧi hik	'to take care of'	ʤaʤ	'hair, fur'
	ʧunʤo˞ik	'magpie'	ʤinʤu	'thorn tree'
c.	ʈʂʰiʈʂ-ik	'to learn'	ɖʐaʈʂ	'spirit beings'
	ʈʂʰãʈʂ-ik	'to pierce'	ʈʂãɖʐa	'pinewood torch'

d. *ʦ...ʈʂ, *ʈʂ...ʦ, *ʧ...ʈʂ, *ʈʂ...ʧ, etc. (no retroflexes with non-retroflexes)

(12) Kalasha: Retroflex consonant harmony between sibilant fricatives

a.	sastir-ik	'to roof a house'	sazu djek	'to have a cold'
	sos	'insides (intestines)'	ispres	'mother-in-law'
b.	ʃiʃoa	'handsome'	ʒoʃi	'Spring festival'
	ʃãʃ	'fishhook'	iʃpaʃur	'father-in-law'
c.	ʂiʂ	'head, top'	ʂuʂ-ik	'to dry'
	ʂiɫeʂ	'glue'	iʂpoʂi	'nephew, niece'

d. *s...ʂ, *ʂ...s, *ʃ...ʂ, *ʂ...ʃ, etc. (no retroflexes with non-retroflexes)

(13) Kalasha: No retroflex consonant harmony between plosives and sibilants

a.	tsaʈẽg-ik	'to move, shake'	ʈrits	'(kind of) bird'
	saʈuk	'apple sauce'	ʈosu djek	'to peck'
	diʈʂ	'period of abstinence'	ʈʂʰet	'cultivated field'
	tuʂ	'straw, chaff'	ʂit	'tight-fitting'
b.	ʧuʈ-ik	'to touch'	ʈõʧuk	'active'
	ʃoʈʰa	'a growth'	ɖʰuʃak	'a dance'

In every respect outlined above, Kalasha and Indus Kohistani are alike. However, they differ in the treatment of some affricate-fricative pairs. Whereas Indus Kohistani enforces retroflex consonant harmony in all affricate-fricative pairs, Kalasha enforces it only in those containing a retroflex affricate in combination with a fricative, as shown in (14), but not in those containing a retroflex fricative in combination with a palatal affricate, as shown in (15).

(14) Kalasha: Retroflex harmony between retroflex affricates and fricatives

a.	ʈʂaʂa	'cottage cheese'	ʈʂaʂku	'evergreen tree'
	zaʈʂgʰur	'half-lame'	ʂanɖʐu-ik	'to wrinkle'

b. *s...ʈʂ, *ʈʂ...s, *ʃ...ʈʂ, *ʈʂ...ʃ, etc. (no retroflexes with non-retroflexes)

(15) Kalasha: No retroflex harmony between retroflex fricatives and palatal affricates

ʧaʂ	'lunch'	ʤoʂ-ik	'to consider'
ʧuʂ-ik	'to suck'	ʤeʂʈ ã˞gu	'thumb, big toe'
ʂaʧ	'temporary shelter'	ʂinʤ-ik	'to win over'

The difference between Kalasha and Indus Kohistani can be seen most clearly by comparing the co-occurrence of palatal and retroflex sibilants in each language. This is presented schematically in (16), where Č represents any palatal affricate, Ç represents any retroflex affricate, Š represents any palatal fricative, and Ṣ any retroflex fricative. A check mark (✓) indicates that the co-occurrence of a pair is preferred, or at least unprohibited. An asterix (*) indicates that the pair is dispreferred or avoided altogether.

(16) Co-occurrence of palatal and retroflex sibilants in Indus Kohistani and Kalasha

	Indus Kohistani				Kalasha			
a.	✓Č-Č	✓Ç-Ç	*Č-Ç	*Ç-Č	✓Č-Č	✓Ç-Ç	*Č-Ç	*Ç-Č
b.	✓Š-Š	✓Ṣ-Ṣ	*Š-Ṣ	*Ṣ-Š	✓Š-Š	✓Ṣ-Ṣ	*Š-Ṣ	*Ṣ-Š
c.	✓Č-Š	✓Ç-Ṣ	*Č-Ṣ	*Ç-Š	✓Č-Š	✓Ç-Ṣ	✓Č-Ṣ	*Ç-Š
	✓Š-Č	✓Ṣ-Ç	*Ṣ-Č	*Š-Ç	✓Š-Č	✓Ṣ-Ç	✓Ṣ-Č	*Š-Ç

Indus Kohistani and Kalasha both enforce agreement for retroflexion when co-occurring sibilants have identical manners, whether the sibilants are both affricates, as in (16)(a), or both fricatives, as in (16)(b). The languages also exhibit similar co-occurrence patterns when the sibilants disagree for manner, as in (16)(c). Both languages have affricate-fricative pairs agreeing in retroflexion or non-retroflexion (i.e., ✓Č-Š, ✓Š-Č, ✓Ç-Ṣ, ✓Ṣ-Ç). In addition, both languages prohibit affricate-fricative pairs that combine retroflex affricates with palatal fricatives (i.e., *Ç-Š, *Š-Ç). The absence of these pairs suggests that they have been subject to retroflex assimilation. However, the languages differ when it comes to affricate-fricative pairs that combine retroflex fricatives with palatal affricates (i.e., Č-Ṣ, Ṣ-Č). These pairs are outlined in (16)(c) to draw attention to them. Whereas Indus Kohistani prohibits pairs of this type, and subjects them to retroflex assimilation, Kalasha does not. This is evident in cognates such as Indus Kohistani /ʈʂoːʂáṽ/ 'to suck (out)' and Kalasha /ʧuʂik/ 'to suck'. Both of these words derive from OIA /ʧuːʂati/ 'sucks' (CDIAL 4898), but whereas Indus Kohistani has subjected the sibilants of this root to retroflex assimilation, Kalasha has not.

It might be possible to interpret the co-occurrence of sibilants in Kalasha as evidence of a further similarity effect; one in which retroflex harmony is sensitive to the distinction between affricates and fricatives within the sibilant class. However, it is not clear that similarity alone can account for the pattern. Not all affricate-fricative pairs co-occur freely in Kalasha. As we have seen, those that combine retroflex affricates with palatal fricatives are clearly avoided, just as they are in Indus Kohistani (i.e., *C̣-Š, *Š-C̣). If retroflex harmony did not operate across the affricate vs. fricative distinction, we would expect these sequences to occur freely. Their absence suggests that they have been subject to assimilation. If so, then at least some cases of consonant harmony in Kalasha are not sensitive to the affricate vs. fricative distinction.

A full account of the Kalasha pattern may need to consider other contributing factors. For instance, palatal affricates may be dispreferred as targets of assimilation, retroflex affricates may be preferred as triggers, or both. There may be historical (or other) factors favouring the retention of palatal affricates in all but the strongest assimilation environments (i.e., in the context of a retroflex segment with identical manner). Alternatively, retroflex affricates may constitute stronger triggers of assimilation than their fricative counterparts, thereby enabling them to trigger assimilation in sibilants of all kinds, while retroflex fricatives are only able to trigger assimilation in other fricatives. The most appropriate explanation for the asymmetry among Kalasha sibilants remains unclear at present. It seems likely that similarity of manner plays at least some role, though it may not tell the whole story.[10] Whatever the case may be, it

---

[10] There is at least one way in which the palatal affricates of Kalasha differ from the retroflex affricates historically. While the retroflex series is an innovation of the Dardic languages, the palatal series is inherited from Old Indo-Aryan (§2.2.3). In Old Indo-Aryan Sanskrit, the palatal affricates patterned phonologically with the non-sibilant stops (i.e., /t, ʈ, ʧ/ vs. /s, ʂ, ʃ/). For instance, they exhibited the same oppositions for voicing and aspiration

is clear that retroflex consonant harmony in Kalasha is sensitive to the sibilant vs. non-sibilant distinction, as it is in Indus Kohistani.

Other similarity effects can also be observed in the Dardic languages. For instance, in addition to retroflex obstruents, Indus Kohistani has a sonorant retroflex flap /ɽ/. This phoneme does not trigger assimilation in obstruents, despite the fact that it derives historically from intervocalic retroflex plosives. Representative examples are listed in (17).

(17) Indus Kohistani: No retroflex harmony between sonorant flaps and plosives

tʌɽʌ̀qʰ	'a blow, knock, stroke'
tʰãːɽã̀ʰ	'police station'
dʰaːɽàʰ	'a cattle raid'
dʰúɽ	'dust'

The retroflex flaps of Indus Kohistani derive historically from the lenition of OIA intervocalic retroflex plosives and nasals. The same diachronic process was taken a step further in Kalasha, where the original retroflex consonant has merged completely with the surrounding vowels. As a result, Kalasha has a full set of retroflex vowels that contrast synchronically with their non-retroflex counterparts (Heegård & Mørch, 2004) (§3.3.2.3). Like the retroflex flaps of

---

as other stops, and they occurred in phonotactic positions typically occupied by other stops, including homorganic nasal + C and fricative + C clusters. Their status as affricates was not phonologically significant. It is tempting to think that Kalasha might preserve the non-sibilant status of palatal affricates and that this might account for their lack of interaction with retroflex fricatives in consonant harmony. Note, however, that palatal affricates do not harmonize with non-sibilant plosives (such as /ʈ/) but do interact with sibilant affricates (such as /ʈʂ/). Thus, it remains unclear whether this (or any other) historical factor can account for the asymmetry in Kalasha.

Indus Kohistani, the retroflex vowels of Kalasha do not serve as triggers of retroflex consonant harmony. They do not trigger regressive assimilation in dental and palatal obstruents the way that other retroflex obstruents do. Representative examples are listed in (18).

(18) Kalasha: No retroflex harmony between retroflex vowels and coronal obstruents

t^{h}ũ˞	‘post, column’	ud^{h}ũ˞	‘dust’
ʦũ˞	‘nil, zero’	ʧu˞i	‘braid, plait’
sũ˞a	‘gold’	ʃa˞	‘reed, arrow’

In summary, we have seen that Kalasha and other Dardic languages, such as Indus Kohistani, provide unambiguous evidence of similarity effects in retroflex consonant harmony. In these languages, retroflexion is contrastive in both sibilant and non-sibilant obstruents. Nevertheless, retroflex consonant harmony applies only to pairs of obstruents that belong to the same manner class, whether both are sibilant or non-sibilant. It does not apply when one is sibilant and the other non-sibilant. Kalasha may show further sensitivity to the affricate vs. fricative distinction within the sibilant class. However, this pattern exhibits some unusual asymmetries that may not be reducible to similarity effects alone. Both languages also show sensitivity to the sonorant vs. obstruent distinction. The retroflex flaps of Indus Kohistani do not trigger assimilation in obstruents, nor do the retroflex vowels of Kalasha. Thus, retroflex consonant harmony in the Dardic languages is highly and unambiguously sensitive to the similarity of participating segments.

4.3.1.3 Transparency and blocking

As discussed in §3.3.4, Dardic languages such as Kalasha and Palula may exhibit blocking effects in retroflex consonant harmony. More specifically, it appears that harmony between plosives may be blocked by intervening sibilants in T-ṢṬ configurations. The limited data bearing on this issue are repeated in (19) below.

(19) Possible examples of blocking in Dardic T-ṢṬ sequences

		Kalasha	Palula	Kohistani	OIA	CDIAL
a.	'to stand'	ʧiʂʈik	—	—	tiʂʈʰati	5837
	'hand span'	ʤiʂ(ʈ)	diʂʈ	ɖíːʈʰi, díʈʰ	diʂʈi	6343
	'female spirit'	ʤeʂʈak	—	—	deːʂʈriː	6556
	'wine'	—	—	diʂʈʌ́ṽ	??	??
b.	'bitter'	ʧiʂʈaka	tríʂʈu	ʧîʈʰi	tr̩ʂʈá-	5938
	'visible; seen'	drẽʂ(ʈ)	dʰriʂʈu	—	dr̩ʂʈá	6518
	'written cure'	draʂʈaw	—	—	??	??

The examples in (19) suggest that, wherever OIA /ʂʈ/ clusters have been preserved, the retroflex plosives in these clusters have not triggered harmony in preceding dental plosives. Rather, Kalasha has palatalized OIA dental plosives in TVṢṬ sequences whenever they occurred before front vowels (e.g., /ʧiʂʈ-/ < /tiʂʈʰ-/). Palatalization entails affrication in Kalasha. It is tempting to think that affrication has bled retroflex consonant harmony in Kalasha by rendering the initial obstruent dissimilar to the triggering retroflex plosive. However, this may not be the case since there are disharmonic word forms even in cases where palatalization/affrication has not occurred (e.g., Palula /diʂʈ/ < OIA /diʂʈi/ in (19)(a) and TrVṢṬ sequences in both Palula and Kalasha in (19)(b)). Taken together, the limited evidence in (19) suggests that retroflex consonant harmony between plosives has been blocked whenever the sibilant in a T-ṢṬ sequence has been preserved, but not when the sibilant has been lost (e.g., Kohistani /ɖíːʈʰⁱ/ ~ /díʈʰ/ < OIA /diʂʈi/ in (19)(a)).

Blocking effects are not generally expected under feature agreement. Thus, the possibility of blocking in Dardic appears to provide a counter-argument to the claim that retroflex consonant harmony in this language group is the product of feature agreement, not feature spreading. However, as pointed out in §3.3.4, the apparent blocking effect in Dardic might not be a true case of blocking, as it is normally understood. Within feature spreading models, blocking arises when the spreading feature encounters an intervening segment that is either incompatible with it or unable to propagate it any further. In the case of Dardic, however, the apparent blocking effect might be regarded as a kind of similarity effect in which assimilation is not blocked in this sense, but simply fails to be triggered in the first place. If long-distance assimilation is triggered only under similarity conditions, and if similarity is evaluated (at least partly) on the basis of acoustic/perceptual properties or features, then the presence of a fricative in the /-ʂʈ-/ cluster might mask or dominate the cues of the plosive to some degree. Thus, Dardic languages might treat /T…ʂʈ/ sequences in a way comparable to /T…ʂ/ sequences or even /T…ʈʂ/ sequences, in which C_2 is an affricate consisting of a phonetic plosive + fricative sequence. In each case, the cues of the fricative element dominate those of any adjacent plosive element thereby rendering the plosive element distinct from simple plosives. As a result, there is no demand for similarity-induced assimilation.[11]

[11] This hypothesis could predict that harmony would not be blocked if the intervening fricative is non-adjacent to the surrounding plosives (T…Ṣ…Ṭ). Unfortunately, there are no roots of this type that might speak to the issue.

In light of the fact that sibilants may interfere with harmony in T-ṢṬ configurations, it is interesting to note that nasals do not block harmony in T-ṆṬ configurations. This may be tied to the fact that nasals have very weak place cues while sibilants have very robust cues (Jun, 2004; Steriade, 2001). Moreover, in the South Asian context, VNC sequences often vary freely with Ṽ^{N}C sequences (or even ṼC sequences), in which the homorganic nasal consonant is realized more as a feature of nasalization on the preceding vowel than as a fully articulated consonant. Both of these facts suggest that a nasal would be less likely than a sibilant to interfere with the cues of the following stop.

Other evidence from Kalasha suggests that the vowels in consonant harmony domains are not targeted or permeated by retroflexion. Proponents of gesture extension models have argued that retroflexion can permeate or spread through intervening vowels with little or no noticeable effect because all features/gestures pertaining to the orientation of the tongue tip are generally irrelevant for vowels (e.g., Gafos, 1999). However, Kalasha is one of those rare languages in which retroflexion *is* relevant for vowels. Recall that Kalasha has a full set of phonemic retroflex vowels (§3.3.2.3). These vowels have developed historically from the coalescence of non-retroflex vowels with intervocalic retroflex consonants (e.g., /kila˞/ 'kind of cheese' < OIA /kilaːʈa-/; /pẽ˞/ 'palm of hand' < OIA /paːɳi-/) (Heegård & Mørch, 2004). Since the feature or gesture that distinguishes retroflex vowels from their non-retroflex counterparts derives historically from consonants, it is reasonable to assume that the retroflex feature/gesture is fundamentally the same for both consonants and vowels in the language. In spite of this, there is no evidence that retroflex vowels participate in retroflex consonant harmony either as triggers, targets or blockers. We have already seen that retroflex vowels do not trigger assimilation in preceding coronal obstruents (§4.3.1.2). There is also evidence to suggest that intervening vowels are not targeted in retroflex consonant harmony. In Trail & Cooper's (1999) Kalasha dictionary, there are no examples of retroflex vowel phonemes occurring in retroflex consonant harmony domains (i.e., between two obstruents that agree in manner and retroflexion). Thus, the vowels in these domains are not perceived as phonemic retroflex vowels by Kalasha speakers. This suggests that vowels do not serve as targets of assimilation even when they stand between the trigger and target of retroflex consonant harmony.

The complete absence of retroflex vowel phonemes in consonant harmony domains could indicate that the contrast between retroflex and non-retroflex vowels is neutralized in this context. However, even if this is the case, it remains significant that the contrast is neutralized in favour of non-retroflex vowels. Under feature spreading, where retroflexion targets or permeates intervening segments, we would expect neutralization to favour retroflex vowels in harmony domains. Thus, the absence of retroflex vowel phonemes in harmony domains suggests that Kalasha vowels are transparent to retroflex consonant harmony.

In sum, there is limited evidence suggesting that retroflex consonant harmony between plosives might be blocked by intervening sibilants in Kalasha and Palula. However, the evidence bearing on this issue is sketchy and the pattern might reflect a similarity effect as opposed to a true blocking effect. Moreover, there is good evidence that Kalasha vowels are neutral and transparent with respect to consonant harmony between obstruents. Intervening vowels in consonant harmony domains are phonologically non-retroflex, despite the fact that Kalasha distinguishes retroflex and non-retroflex vowel phonemes in other contexts.

In this section we have seen that retroflex consonant harmony in Kalasha and other Dardic languages is characterized by typological properties that are consistent with feature agreement. These include: (i) regressive directionality; (ii) sensitivity to the similarity of participating segments (that cannot be reduced to the influence of contrast); and (iii) transparency of intervening segments. Of these properties, the similarity effects are the most compelling. The significance of the similarity effects and the regressive directionality become all the more evident when the pattern of consonant harmony in Kalasha is compared with patterns of local retroflex assimilation in the language. §4.3.2 examines assimilatory co-occurrence restrictions on adjacent coronal consonants in Kalasha and §4.3.3 reviews the pattern of retroflex vowel-consonant harmony in the language. Both of these patterns exhibit a very different set of typological properties; one that is more in keeping with feature spreading.

4.3.2 Local retroflex consonant assimilation

The system of retroflex consonant harmony in Kalasha stands in sharp contrast with the assimilatory co-occurrence restriction on adjacent coronal consonants in the language. In Kalasha, morpheme-internal consonant clusters consisting of two coronal obstruents are restricted to sequences of fricative + plosive. These clusters show agreement for retroflexion or non-retroflexion *without regard for similarity* along the sibilant/non-sibilant dimension. Representative examples are shown in (20).

(20) Kalasha: Assimilatory co-occurrence restrictions in fricative + plosive clusters

a.	post	'skin'	iston	'udder'
	asta	'also, too'	nast	'nose'
b.	paʃt	'ribs'	ʃuruʃtju	'thoroughness'
	piʃtjak	'behind'	paʃtari	'power'

c. piʂʈ 'upper back' iʂʈep karik 'to suffocate'

aʂʈ(a) 'eight' uʂʈ 'lip'

d. *sʈ, *ʂt, *ʃʈ (no retroflexes with non-retroflexes)

Kalasha overwhelmingly prefers fricative + plosive clusters consisting of homorganic dental consonants (i.e., /st/), as in (20)(a), or retroflex consonants (i.e., /ʂʈ/), as in (20)(c). Palatal fricatives and dental plosives can also co-occur in some clusters (i.e., /ʃt(j)/, as shown in (20)(b). However, examples of this kind are less frequent. Moreover, the dental plosives in these sequences are often followed by a palatal glide or high front vowel, suggesting that they might be somewhat palatalized. Whatever the case may be with palatalization, it is clear that the language systematically avoids fricative + plosive clusters that disagree in retroflexion, including */sʈ/, */ʂt/, and */ʃʈ/ (cf. Morgenstierne 1973, p. 192). The assimilatory co-occurrence restriction on adjacent obstruents in (20) stands in sharp contrast to the restriction on non-adjacent obstruents in (10)–(13). Unlike retroflex consonant harmony between non-adjacent obstruents, local assimilation in Kalasha is not sensitive to similarity along the sibilant/non-sibilant dimension.[12]

Kalasha also has morpheme-internal consonant clusters consisting of nasal + stop. Again, if both consonants in the cluster are coronal they are invariably homorganic. This is not

---

[12] Indus Kohistani has not preserved OIA fricative + stop clusters (cf. §3.3.4). Thus, it is not possible to compare the local assimilation of obstruents with the pattern of retroflex consonant harmony in that language. However, a few examples of fricative + plosive clusters can be found in loanwords and these are always homorganic. For instance, /st/ clusters can be found in a few Persio-Arabic loans (e.g., /dʌstíː/ 'immediately'; cf. Persian /dastī/). A single example of /ʂʈ/ is found in the word /diʂʈʎ́ṽ/ 'wine', which may be a loanword from another language of the region (cf. Burushaski /diʂáo/ 'eingekochter Traubensaft [boiled grape juice]' in Berger, 1998b, p. 120).

always shown in phonemic transcriptions because palatal and retroflex nasals are often treated as allophones of dental /n/.[13] Nevertheless, at a phonetic level, coronal nasal + stop sequences are always homorganic. Thus, we find dental sequences such as [nd] in words like /kanda/ ‘wild almond tree’, palatal sequences such as [ɲʤ] in words like /menʤ/ ‘cloud’, and retroflex sequences such as [ɳɖ] in words like /gonɖ/ ‘pole’. Coronal nasal + stop sequences never disagree phonetically in retroflexion (i.e., *[nɖ], *[ɲɖ], *[ɳd], *[ɳʤ], etc.) (cf. Mørch & Heegård, 1997, p. 51; Heegård & Mørch, 2004, pp. 65–67; Morgenstierne, 1973, pp. 190, 192). The same is true of nasal + stop clusters in Indus Kohistani (e.g., /zʌ́nd̥/ ‘alive’, but /zʰúɳɖ̊/ ‘a branch of a holm oak’). Thus, local retroflex assimilation in Kalasha and Indus Kohistani is not sensitive to similarity in terms of the sonorant vs. obstruent distinction.

There is also some evidence that local assimilation can hold between obstruents and glides across morpheme boundaries, resulting in de-retroflexion. For example, /ʧot-jak/ ‘tattoo marks’ may be derived historically, if not synchronically, from /ʧoʈ/ ‘design, pattern’ followed by the diminutive suffix /-jak/. If so, then the palatal glide of the suffix has induced de-retroflexion of the stem-final retroflex plosive. A similar relation probably holds between lexical pairs such as /piʂʈ/ ‘upper back’ in (20)(c) and /piʃt-jak/ ‘behind’ in (20)(b).[14]

---

[13] The transcription in Trail & Cooper’s (1999) dictionary of Kalasha, which is the primary source of data for the current study, does not distinguish dental and retroflex nasals but treats them both as allophones of /n/. Other studies treat the retroflex nasal as having a marginal phonemic status outside of homorganic NC clusters (Morgenstierne, 1973; Mørch & Heegård, 1997; Heegård & Mørch, 2004).

[14] Trail & Cooper (1999) list /ʧotjak/ ‘tattoo marks’ and /piʃtjak/ ‘behind’ as independent lexical items containing the dental phoneme /t/. However, both items are cross-referenced to roots in which dental /t/ corresponds to retroflex /ʈ/: /ʧoʈ/ and /piʂʈ/, respectively. This suggests that de-retroflexion has been triggered by the diminuitive suffix /-jak/, but it is not clear whether this process is productive synchronically or whether it reflects a diachronic

The direction of assimilation in morpheme-internal clusters is difficult to assess because we are dealing with a static co-occurrence restriction. Nasal place assimilation tends to be regressive cross-linguistically. Thus, assimilation is arguably regressive in the case of nasal + stop sequences. This is consistent with the phonemic forms assumed in Trail & Cooper's (1999) transcription of Kalasha, in which phonetic retroflex and palatal nasals are treated as /n/ before retroflex and palatal stops. This phonemic interpretation assumes that retroflex and palatal nasals are the product of regressive place assimilation.

The direction of assimilation in fricative + plosive sequences is more difficult to determine because the co-occurrence pattern has a long history that goes back to the earliest OIA period. There is evidence that this pattern was originally progressive. Retroflex fricatives were among the earliest retroflex phonemes in Indo-Aryan. One of the factors that introduced retroflex plosives was progressive assimilation induced by preceding retroflex fricatives (i.e., *ʂt → ʂʈ). (Misra B. G., 1967, pp. 28–29, 63ff; Bhat D. N., 1973, p. 33; Hamp, 1996) (§2.2.3). However, in OIA Sanskrit there is evidence of both progressive assimilation (e.g., /iʂ + ta-/ → [iʂʈa-] 'desired') and regressive assimilation (e.g., /kuʈʰaːrais ʈaŋkais ʧa/ → [kuʈʰaːraiʂ ʈaŋkaiʃ ʧa] 'with axes and crowbars').[15] Thus, from a diachronic perspective, assimilation between obstruents can probably be regarded as bi-directional.

---

development. If it is productive synchronically, then these words might be represented phonemically as /ʧoʈ-jak/ and /piʃʈ-jak/, with retroflex /ʈ/ in place of dental /t/.

[15] The Sanskrit examples are adapted from Stenzler (1997, pp. 9, 11). The final /s/ in words such as /kuʈʰaːrais/ reflects orthographic *visarga*, i.e., a voiceless glottal fricative [h] (represented as [ḥ] in traditional roman transliterations). This segment is regarded as an allophone of /s/ in final position (Masica, 1991, p. 161).

In sum, assimilatory co-occurrence restrictions on adjacent coronal consonants in Kalasha exhibit typological properties consistent with local feature spreading. Assimilation is not sensitive to the similarity of participating consonants. All coronal consonant clusters show agreement for retroflexion or non-retroflexion regardless of whether they consist of fricative + plosive or nasal + stop. From a synchronic perspective, the pattern is manifested as a static morpheme structure constraint with no clear direction of assimilation. From a diachronic perspective, however, there is evidence of both progressive and regressive assimilation. Thus, local retroflex assimilation between consonants shows no inherent directional bias and no sensitivity to the similarity of participating segments.

4.3.3 Retroflex vowel-consonant harmony

As discussed in §3.3.2.3, Kalasha exhibits contrastive retroflexion in vowels. All vowels, whether oral or nasal, have phonemic retroflex and non-retroflex counterparts. While these vowels do not participate in the pattern of retroflex consonant harmony involving obstruents, they do participate in another pattern of retroflex assimilation; one that is distinct from retroflex consonant harmony. The vocalic pattern can be described as retroflex vowel or vowel-consonant harmony. Representative examples are repeated in (21) below.

(21) Retroflex vowel and vowel-consonant harmony in Kalasha (Heegård & Mørch, 2004)

a.	/pr̃˞ik/	[pr̃˞ik] ~ [pr̃˞i˞k]	‘to squeeze’
	/ʧa˞ha˞ka/	[ʧa˞ha˞ka] ~ [ʧa˞ha˞ka˞]	‘maize bread’
	/a˞ngu/	[a˞ŋgu] ~ [a˞ŋgu˞]	‘finger’
	/sirã˞/	[sirã˞] ~ [si˞rã˞]	‘wind’
b.	/a˞in/	[a˞in] ~ [a˞i˞ɳ]	‘millet’

The examples in (21) demonstrate that retroflexion can assimilate (optionally) from one vowel to another. This pattern appears to respect strict locality. No vowels are skipped and even though some intervening consonants appear to be transparent, they are most likely permeated by the retroflex feature or gesture. Evidence for this comes from examples like (21)(b), where assimilation targets a coronal nasal /n/ that lies in its path. Not only does this pattern of assimilation respect strict locality, it also shows no evidence of similarity effects. All vowels participate in the pattern regardless of tongue height, backness, lip rounding or nasalization. Moreover, as the example in (21)(b) shows, consonants can also serve as targets.

Assimilation in (21) is predominantly progressive, although regressive assimilation also appears to be possible, as suggested by /sirã˞/ [sirã˞] ~ [si˞rã˞] 'wind'. Thus, with respect to directionality, retroflex vowel-consonant harmony in Kalasha is predominantly progressive but potentially bi-directional.

In sum, retroflex vowel-consonant harmony in Kalasha exhibits typological traits in keeping with local feature spreading. It appears to obey strict locality, exhibits progressive or bi-directional assimilation, and shows no obvious similarity effects.

The evidence from Kalasha provides compelling support for the distinction between feature agreement and feature spreading. Kalasha exhibits three identifiable patterns of retroflex assimilation: retroflex consonant harmony, local retroflex consonant assimilation, and retroflex vowel-consonant harmony. The first of these exhibits properties consistent with feature agreement while the others exhibit properties consistent with feature spreading. In the case of retroflex consonant harmony, assimilation is: (i) regressive; (ii) sensitive to similarity; and (iii) not dependent on strict segmental adjacency so that intervening segments can be skipped. These are precisely the properties expected under feature agreement. In the case of local consonant assimilation and vowel-consonant harmony, assimilation is: (i) progressive or bidirectional; (ii) not sensitive to similarity; and (iii) subject to strict locality so that intervening segments are not skipped. These are precisely the properties expected under feature spreading.

4.4 Summary and conclusions

The hypothesis that assimilation is driven by two independent mechanisms, feature spreading and feature agreement, is based largely on the observation that consonant harmony systems exhibit unique typological properties that set them apart from other patterns of assimilation. This typological distinction is unexpected if all assimilation is the product of local feature spreading. However, it receives a natural explanation if some types of assimilation are driven by local spreading while others are driven by long-distance agreement.

Each mechanism of assimilation has its own functional grounding and associated typological properties. Feature spreading is typically grounded in low-level articulatory factors such as co-articulation and articulatory simplification. It operates locally over spans of adjacent segments, shows a general disregard for similarity, and no inherent directional bias. Feature agreement is grounded in higher-level cognitive functions associated with speech planning. It operates over non-adjacent segments, is highly sensitive to similarity, and exhibits a bias for anticipatory/regressive directionality. Some patterns of apparent long-distance retroflex assimilation in South Asia exhibit properties consistent with feature spreading, but most exhibit properties consistent with feature agreement. Patterns of both types co-exist in Kalasha, which provides strong evidence for the distinction between the two.

It might be helpful to consider both types of retroflex assimilation in terms of the functional requirements of serial-order production introduced earlier: (i) a 'turn-on' function, in which the system must activate the present; (ii) a 'turn-off' function, in which the system must deactivate the past; and (ii) a 'prime' function, in which the system must prepare to activate the future. A failure to execute the 'turn-off' function will result in the progressive spread of a feature/gesture over a contiguous span of segments, without regard for similarity, until the articulatory requirements of another segment demand a change. This is essentially the explanation of Sanskrit n-retroflexion offered by Whitney (1993 [1889]):

> *We may thus figure to ourselves the rationale of the process: in the marked proclivity of the language toward lingual* [=retroflex] *utterance, especially of the nasal, the tip of the tongue, when once reverted into the loose lingual position by the utterance of a non-contact lingual element* [=retroflex continuant]*, tends to hang there and make its next nasal contact in that position; and does so, unless the proclivity is satisfied by the utterance of a lingual mute* [=retroflex stop]*, or the organ is thrown out of adjustment by the utterance of an element which causes it to assume a different posture (Whitney, 1993 [1889], p. 65).*

Whitney's explanation of Sanskrit n-retroflexion assumes that the retroflex feature/gesture is sustained over a contiguous span of segments until it meets a segment that is either incompatible with it or fails to propagate it further. This explanation of Sanskrit is also advocated by Allen (1951), Gafos (1999), Ní Chiosáin & Padgett (1997), and Hansson (2010). The present study assumes that this explanation is essentially correct for Sanskrit n-retroflexion

(§3.2.1.1), and that a similar explanation can also be extended to retroflex vowel-consonant harmony in Kalasha (§3.3.2.3), progressive retroflex assimilation in Sherpa (§3.5), and possibly alternations in the non-past suffix of Burushaski (§3.3.5.2), all of which exhibit similar typological properties.

In contrast with these, most other cases of long-distance retroflex assimilation reviewed in Chapter 3 exhibit properties in keeping with feature agreement. Agreement arises not from the failure to execute the 'turn off' function of serial-order production, but rather, from interference between the 'turn on' and 'prime' functions, which are largely concurrent. If the 'prime function' is preparing to activate an upcoming segment while the 'turn on' function is engaged in activating a highly similar segment, then there is the potential for similarity-induced anticipatory intrusions. Over time, agreement of this type might lead to phonetic variation and ultimately to grammaticalized morpheme structure restrictions favouring 'harmonic' forms in the lexicon. All cases of long-distance regressive retroflex assimilation in South Asia that exhibit similarity effects are best understood in this way.

The weight of evidence from retroflex assimilation in Kalasha and other South Asian languages supports the typological distinction between feature agreement and feature spreading, and the conclusion that retroflex consonant harmony in South Asia is the product of the former, not the later. The following chapter introduces the Agreement by Correspondence (ABC) model of feature agreement and demonstrates how it can be extended to retroflex consonant harmony in South Asian languages.

Chapter 5
An ABC account of retroflex agreement

The preceding chapter reviewed typological evidence from South Asian languages supporting the hypothesis that feature agreement and feature spreading constitute two independent mechanisms of assimilation, and that retroflex consonant harmony in South Asia is the product of agreement, not spreading. This chapter provides a formal account of retroflex consonant harmony as agreement, within the Agreement by Correspondence (ABC) model proposed by Rose & Walker (2004) and Hansson (2001; 2010).[1]

The chapter is structured as follows. Section §5.1 briefly outlines the phonological features assumed for the purpose of the analysis. The ABC model and associated constraints are introduced in §5.2. Section §5.3 demonstrates how the model can account for similarity-induced long-distance retroflex assimilation in South Asian languages. A few outstanding issues are discussed in §5.4 and some concluding remarks are offered in §5.5.

5.1 Phonological features

A formal account of feature agreement presupposes some set of phonological features. Unfortunately, there is no universally recognized set of features. Proposals for features representing retroflexion range from a simple monovalent feature such as [retroflex] (Ní Chiosáin & Padgett, 1997), or its bivalent counterpart [±retroflex] (Ohala, 1983), to more

---

[1] A formal account of local retroflex feature spreading is beyond the scope of the present study and will not be pursued here. For possible approaches to retroflex feature spreading see Ní Chiosáin & Padgett (1997) and Gafos (1999).

complex representations involving multiple articulatory features such as [−anterior, −distributed] (Chomsky & Halle, 1968) and [−distributed, +back] (Gnanadesikan, 1994), or acoustic features such as [−grave, −sharp, +flat] (Hamilton, 1996), just to name a few. The question of phonological features is orthogonal to the issue at hand, which concerns the modelling of feature agreement, regardless of the features involved. For convenience, I assume the binary features of Chomsky & Halle (1968) supplemented with monovalent major articulator features, such as [CORONAL], as introduced by Sagey (1986). This combination is often employed in the literature and is perhaps the closest thing to a “standard” feature set in generative phonology (e.g., Hall, 2007).

The most relevant features assumed for the current analysis are summarized in Table 32, Table 33 and Table 34. In these tables, a check mark (✓) indicates the applicability of a monovalent feature, while plus (+) and minus signs (−) denote the positive and negative values of binary features.

Table 32 Place features

		p	t̪	t̺	ʈ	tʃ	k
[LABIAL]	([LAB])	✓					
[DORSAL]	([DOR])						✓
[CORONAL]	([COR])		✓	✓	✓	✓	
[anterior]	([ant])		+	+	−	−	
[distributed]	([dist])		+	−	−	+	

Table 33 Manner and stricture features

		t, ʈ	ʦ, ʈʂ, ʧ	s, ʂ, ʃ	n, ɳ	l, ɭ	r, ɽ
[sonorant]	([son])	−	−	−	+	+	+
[continuant]	([cont])	−	−	+	−	−	+
[strident]	([strid])	−	+	+	(−)	(−)	(−)
[nasal]	([nas])	−	−	−	+	−	−
[lateral]	([lat])	−	−	−	−	+	−

Table 34 Laryngeal features

		t, ʈ	tʰ, ʈʰ	d, ɖ	dʰ, ɖʰ
[voice]	([voi])	−	−	+	+
[spread glottis]	([s.g.])	−	+	−	+

In what follows, I assume the feature specifications outlined in the tables above. Nothing critical hinges on the choice of this particular set of features over any other. The account sketched here could be easily adapted to accommodate a different set of features.[2]

[2] The choice of features does have implications for a formal analysis to the extent that different features predict different natural classes, which in turn have a bearing on predictions concerning the relative similarity of segments. For instance, the feature [−anterior] predicts that retroflex and palatal segments constitute a natural class distinct from dental segments. This implies that retroflex and palatal segments might be deemed more similar to one another than to dentals, and, therefore, that they might be more likely to interact in long-distance assimilation. This may or may not be a desirable prediction. Elsewhere, the feature [±anterior] has been called into question for predicting unattested natural classes (e.g., Diffloth, 1975; Gnanadesikan, 1994; Arsenault, 2009b). Any undesirable predictions associated with [±anterior] might be avoided by simply adopting a different feature system, for instance, one that makes use of monovalent features such as [retroflex] and [palatal]. This does not affect the essence of the ABC analysis, which is concerned with modeling how similarity serves as a bridge to long-distance agreement. Questions concerning the nature of similarity and how similarity is evaluated are somewhat orthogonal to this and are taken up again in Chapter 6.

5.2 The ABC model

The ‘Agreement By Correspondence’ (ABC) model of consonant harmony assumes the basic framework of Optimality Theory (OT) (Prince & Smolensky, 2004 [1993]). The ABC model has been elaborated most extensively by Rose & Walker (2004) and Hansson (2001; 2010), both building on earlier work by Walker (2000a; 2000b; 2001). All of these accounts assume a common architecture but differ in some details concerning the precise formulation of constraints. The account presented here draws primarily on Rose & Walker (2004).

The ABC model is built on two basic premises. The first asserts that similarity is a source of correspondence between segments in output strings. The second asserts that long-distance agreement is the product of constraints that enforce feature matching between correspondent segments. Thus, at the heart of the ABC model are two constraint families: (i) the CORR-C⇔C constraints that demand correspondence relations between similar segments in output strings (§5.2.1); and (ii) the IDENT-CC constraints that demand featural agreement between segments that stand in a correspondence relation in the output (§5.2.2). These constraints compete with standard OT constraints of the IDENT-IO family that demand faithfulness to input feature specifications (§5.2.3). The basic architecture of the model is represented schematically in (1), and the various constraint families are described below in §5.2.1–§5.2.3.

(1) Consonantal correspondence model (adapted from Rose & Walker, 2004, p. 492)

Input	/ t a ɖ /	
	⇳	IO Faithfulness (IDENT-IO)
Output	[t a ɖ]	
	⬑ ⬏	CC Faithfulness (IDENT-CC)

5.2.1 CORR-C↔C constraints

The first constraint family central to the ABC model is the class of CORR-C↔C constraints responsible for output-output correspondence relations. Rose & Walker (2004) outline the CORR-C↔C constraint schema as shown in (2).

(2) CORR-C↔C: Let S be an output string of segments and let C_i, C_j be segments that share a specified set of features F. If C_i, $C_j \in$ S, then C_i is in a relation with C_j; that is, C_i and C_j are correspondents of one another (Rose & Walker, 2004, p. 491).

CORR-C↔C constraints enforce correspondence relations between consonants in the output string based on their similarity to each other. For instance, a constraint such as CORR-T↔Ṭ requires correspondence between segments that *differ at most* in minor coronal place features such as [±ant, ±dist]. This means that corresponding consonants must agree in all features apart from these, including major place features (e.g., [COR]), manner/stricture features (e.g., [±son, ±strid, ±cont]), and laryngeal features (e.g., [±voi, ±s.g.]). CORR-T↔Ṭ demands that corresponding segments must be at least as similar as /t/ and /ʈ/, though they may also be *more* similar than this. That is, the constraint demands correspondence between segments differing in both [±ant, ±dist] (e.g., /t…ʈ/, /ʦ…ʈʂ/, /s…ʂ/, etc.), as well as those that differ only in [±ant] (e.g., /ʦ…ʧ/, /s…ʃ/, etc.), only in [±dist] (e.g., /ʧ…ʈʂ/, /ʃ…ʂ/, etc.), or neither (i.e., identical pairs such as /t…t/, /ʈ…ʈ/, /ʧ…ʧ/, /ʂ…ʂ/, etc.). The CORR-C↔C constraints required to account for retroflex consonant harmony in Dardic languages such as Kalasha and Indus Kohistani are described in (3).

(3) CORR-C↔C constraints required for retroflex consonant harmony in Dardic languages

a. CORR-T↔Ḍʰ Any two segments with identical specifications for [COR], [±son], [±strid], and [±cont] are correspondents of one another; i.e., all segments differing at most in minor coronal place features, ([±ant], [±dist]) and laryngeal features ([±voi], [±s.g.]).

b. CORR-TS↔Ẓʰ Any two segments with identical specifications for [COR], [±son] and [±strid] are correspondents of one another; i.e., all segments differing at most in minor coronal place features ([±ant], [±dist]), laryngeal features ([±voi], [±s.g.]), and continuancy ([±cont]).

c. CORR-T↔J̣ʰ Any two segments with identical specifications for [COR], [±son] and [±cont] are correspondents of one another; i.e., all segments differing at most in minor coronal place features ([±ant], [±dist]), laryngeal features ([±voi], [±s.g.]), and stridency ([±strid]).

d. CORR-T↔Ẓʰ Any two segments with identical specifications for [COR] and [±son] are correspondents of one another; i.e., all segments differing at most in minor coronal place features ([±ant], [±dist]), laryngeal features ([±voi], [±s.g]), continuancy ([±cont]), and stridency ([±strid]).

e. CORR-T↔Ṛ^{h} Any two segments with identical specifications for [COR] are correspondents of one another; i.e., all segments differing at most in minor coronal place features ([±ant], [±dist]), laryngeal features ([±voi], [±s.g]), continuancy ([±cont]), stridency ([±strid]), and sonorancy ([±son]).

The ranking of CORR-C↔C constraints is partially determined by inherent subset-superset relations. For instance, CORR-T↔Ḍ^{h} outranks CORR-T↔J̣^{h} because the class of segments specified by the former constraint constitutes a proper subset of that specified by the latter constraint. CORR-T↔Ḍ^{h} demands correspondence between all coronal segments differing at most in laryngeal features and minor place, whereas CORR-T↔J̣^{h} demands correspondence between the same set of segments plus those that disagree along the sibilant vs. non-sibilant dimension (i.e., [±strid]). CORR-T↔Ḍ^{h} represents a greater degree of similarity than CORR-T↔J̣^{h}. The ranking of CORR-T↔Ḍ^{h} over CORR-T↔J̣^{h} reflects the fact that the pressure for correspondence increases along with increased similarity, and the fact that any language enforcing correspondence between segments with a low degree of similarity will also enforce it between segments with a higher degree of similarity. On the basis of subset-superset relations, the CORR-C↔C constraints in (3) are ranked as shown in (4).

(4) Similarity-based correspondence hierarchy

CORR-T↔Ḍ^{h}	》	CORR-TS↔Ẓ^{h}	,	CORR-T↔J̣^{h}	》	CORR-T↔Ẓ^{h}	》	CORR-T↔Ṛ^{h}
‘same manner’		‘same stridency’		‘same continuancy’		‘same sonorancy’		‘all coronals’

The hierarchy in (4) correctly predicts that correspondence between segments with a lesser degree of similarity entails correspondence between segments with a greater degree of similarity. Implicational relations and fixed rankings hold between CORR-C↔C constraints whenever the set of features specified in one constraint form a proper subset of those specified in another. However, where feature sets are not in a subset-superset relation, then the relative ranking of constraints may be subject to language-specific variation. For instance, CORR-TS↔Ẓ^h and CORR-T↔J̣^{h} have no inherent ranking in (4) because the class of segments specified by one does not constitute a subset of the other. In cases like this, where constraint ranking cannot be determined on the basis of subset-superset relations, Rose & Walker assume that the relative ranking of constraints must be determined on a language-specific basis or possibly on the basis of some intrinsic universal similarity metric, such as the Natural Classes Similarity Metric of Frisch, Pierrehumbert, & Broe (2004) (to be discussed in §6.1).[3]

5.2.2 IDENT-CC constraints

The second family of constraints central to the ABC model is the class of IDENT-CC constraints. CORR-C↔C constraints establish correspondence relations between consonants in

---

[3] CORR-C↔C constraints might also be parameterized according to the proximity of interacting consonants. For instance, Hansson (2010) proposes constraints of the following types: (i) CORR-$[F]_{CC}$ enforces correspondence between string-adjacent consonants disagreeing at most in the feature [F]; (ii) CORR-$[F]_{C\text{-}V\text{-}C}$ enforces correspondence between transvocalic consonants; and (iii) CORR-$[F]_{C\text{-}\infty\text{-}C}$ enforces correspondence between consonants separated by any number of intervening segments. Correspondence over longer domains entails correspondence over shorter domains and correspondence over shorter domains universally outranks correspondence over longer domains (i.e., CORR-$[F]_{CC}$ ⟫ CORR-$[F]_{C\text{-}V\text{-}C}$ ⟫ CORR-$[F]_{C\text{-}\infty\text{-}C}$). This reflects the fact that demand for correspondence increases along with increased proximity between interacting segments and the fact that any language enforcing correspondence over longer domains will also enforce it over shorter domains. In place of parameterized CORR-C↔C constraints, Rose & Walker (2004) assume a single PROXIMITY constraint requiring correspondent segments to be located in adjacent syllables.

the output string but do not enforce agreement/assimilation between those consonants. The job of enforcing agreement under correspondence falls to the IDENT-CC constraints. These constraints are much like the familiar IDENT-IO constraints of standard OT, except that they refer to output-output relations instead of input-output relations. For instance, the constraint IDENT-CC[−dist] enforces agreement for the feature [−dist] between correspondent segments in the output string in much the same way that IDENT-IO[−dist] enforces agreement for the same feature between input segments and their output correspondents.

Where directional asymmetries are observed, Rose & Walker (2004) encode directionality on the IDENT-CC constraints. Thus, IDENT-CC constraints can be subdivided into IDENT-C_RC_L constraints, which evaluate featural faithfulness from right to left, and IDENT-C_LC_R constraints, which evaluate featural faithfulness from left to right. The directional bias encoded on IDENT-CC constraints is similar to the directional bias inherent in other standard OT faithfulness constraints. For instance, IDENT-IO and MAX-IO constraints are also inherently directional in the sense that they evaluate the output relative to the input, but not *vice versa*. Conversely, IDENT-OI and DEP-IO constraints evaluate the input relative to the output.

The IDENT-CC constraints required for an account of retroflex consonant harmony are described in (5).

(5) IDENT-CC constraints enforcing regressive assimilation of retroflex features

a. ID-C_RC_L[−dist] Let C_L be a segment in the output and C_R be any correspondent of C_L such that C_R follows C_L in the sequence of segments in the output ($C_R > C_L$). If C_R is [−dist], then C_L is [−dist].

b. ID-$C_R C_L$[−ant] Let C_L be a segment in the output and C_R be any correspondent of C_L such that C_R follows C_L in the sequence of segments in the output ($C_R > C_L$). If C_R is [−ant], then C_L is [−ant].

The IDENT-CC constraints in (5) enforce regressive assimilation of coronal place features between correspondent segments in the output string. In an output string of the type $C_1 \ldots C_2$, where C_1 and C_2 stand in a correspondence relation, if C_2 is [−dist, −ant], then the constraints in (5) will demand that C_1 is also [−dist, −ant]. However, if C_1 is [−dist, −ant], then the constraints in (5) will place no demands on C_2 because they only evaluate the faithfulness of correspondents to the left (i.e., preceding correspondents).

5.2.3 IDENT-IO constraints

The CORR-CC and IDENT-CC constraints outlined above interact with IDENT-IO constraints, which demand faithfulness to input features on correspondent output segments. The IDENT-IO constraints that are relevant for the analysis of retroflex consonant harmony in Dardic languages are described in (6) below.

(6) IDENT-IO constraints enforcing faithfulness to input coronal place features

a. ID-IO[±dist] Let C_i be a segment in the input and C_j be any correspondent of C_i in the output. If C_i is [αdist], then C_j is [αdist].

b. ID-IO[±ant] Let C_i be a segment in the input and C_j be any correspondent of C_i in the output. If C_i is [αant], then C_j is [αant].

The constraints in (6) can be broken down into value-specific counterparts. For instance, IDENT-IO[±dist] can be broken down into two constraints: IDENT-IO[+dist] and IDENT-IO[−dist]. Moreover, each value-specific counterpart can be independently ranked in order to reflect the fact that languages can favour faithfulness to one value of a given feature over another.

In summary, the ABC model posits two constraint families: (i) the CORR-C↔C constraints, which demand correspondence relations between similar segments in output strings; and (ii) the IDENT-CC constraints, which demand feature agreement between output segments that stand in a correspondence relation. The following sections demonstrate how the interaction of these constraints, together with standard Faithfulness constraints of the IDENT-IO family, can account for similarity-sensitive retroflex consonant harmony of the type found in South Asian languages, including Dardic languages such as Indus Kohistani and Kalasha.

5.3 Deriving similarity effects in retroflex agreement

Retroflex consonant harmony in South Asia is sensitive to the similarity of interacting segments. In the ABC model, similarity effects are derived through the interaction of CORR-C↔C, IDENT-CC and IDENT-IO constraints. This section demonstrates that an appropriate ranking of these constraints can account for the similarity effects attested in the retroflex consonant harmony systems of South Asian languages.

OT analyses typically operate on the premise that the constraints of the grammar and their ranking should be capable of deriving well-formed outputs regardless of the input (i.e., the principle known as 'Richness of the Base'). In the case of retroflex consonant harmony in South Asia, there is good evidence to support some input-output mappings, but not others. For instance, there is evidence that disharmonic C_1-C_2 input configurations are subject to regressive retroflex assimilation when C_2 is retroflex and C_1 is dental or palatal (e.g., /t…ʈ/ → [ʈ…ʈ]; /ʃ…ʂ/ → [ʂ…ʂ]; etc.). In some cases, there is also a degree of evidence supporting regressive palatal assimilation when C_2 is palatal and C_1 is dental (e.g., /s…ʃ/ → [ʃ…ʃ]; etc.). However, the appropriate output mapping for hypothetical input configurations in which C_1 is retroflex and C_2 is dental or palatal (e.g., /ʈ…t/, /ʂ…ʃ/, etc.), or in which C_1 is palatal and C_2 is dental (e.g., /ʃ…s/, etc.), remains highly speculative. Forms of this type are typically avoided in South Asian languages but there is little or no evidence to indicate what kind of output these hypothetical inputs would map to, should they arise; they might surface faithfully (e.g., /ʈ…t/ → [ʈ…t]), they might be subject to progressive retroflex assimilation (e.g., /ʈ…t/ → [ʈ…ʈ]) or they might be subject to regressive de-retroflexion (e.g., /ʈ…t/ → [t…t]). The account presented in this section focuses on deriving similarity effects in those input-output mappings for which there is relatively unambiguous support. Discussion of other hypothetical input-output mappings is deferred until §5.4.1.

As a general rule, retroflex consonant harmony in South Asia holds between consonants that agree in manner/stricture features but not in laryngeal features. By hypothesis, a correspondence relation must hold between consonants that differ at most in laryngeal features and coronal place features, but not between consonants that differ in manner/stricture features. The relevant correspondence constraint is CORR-T↔Ḍ^{h}. This constraint requires correspondence between output segments that differ at most in laryngeal features and coronal place of articulation, but agree in all other features including those pertaining to manner/stricture.

The constraints that drive retroflex consonant harmony are IDENT-C_RC_L[−ant] and IDENT-C_RC_L[−dist], which demand agreement for retroflex features between correspondent consonants. In the general case, retroflex consonant harmony is enforced when these constraints, together with CORR-T⇔Ḍ^h, outrank IDENT-IO[+ant] and IDENT-IO[+dist], which demand faithfulness to non-retroflex coronal features in the input.

The most basic similarity effect evident in the current study concerns the distinction between obstruents and sonorants. In every case reviewed in Chapter 3, retroflex consonant harmony holds between co-occurring coronal plosives within a root. Wherever retroflexion is contrastive within the class of sonorants, the retroflex sonorants fail to trigger harmony in plosives (or any other obstruents) and also fail to serve as targets of harmony. Representative examples from Malto (Dravidian) and Panjabi (Indo-Aryan) are listed in (7) and (8).

(7) Retroflex consonant harmony in Malto (data from Mahapatra, 1979, 1987)

a. Retroflex harmony between plosives

ʈeʈu	'hand'	< *t…ʈ
ʈuːɖ	'tiger'	< *t…ɖ
ɖeʈa	'corn cob'	< *d…ʈ

b. No retroflex harmony between sonorants and plosives

taɽe	'grinding stone'
daɽe	'animal for sacrifice'

c. Retroflex harmony does not target (initial) sonorants[4]

nuɖ-	'to hide'	noːɽ	'to wash'
laʈa	'gum resin'	loɽa	'a stone to grind spices'
roːɖo	'cork tree'	raːɽe	'enemy'

(8) Retroflex consonant harmony in Panjabi (data from Jain, 1934; Goswami, 2000; Turner, 1962–1966)

a. Retroflex harmony between plosives

ʈaʈʈiː	'screen'	< *t…ʈ
ʈeɖʰaː	'crooked'	< *t…ɖʰ
ɖaʈʈaː	'plug'	< *d…ʈ

b. No retroflex harmony between sonorants and plosives

taːɽ	'palm tree'
toːɽ-naː	'to break'
taːɳ	'tune'
daːɳ	'gift, charity'

---

[4] Based on the case studies in Chapter 3, it is abundantly clear that sonorants are not targeted for retroflex consonant harmony in word-initial C_1VNC_2 domains. In most cases, however, co-occurrence restrictions were not examined systematically outside of these domains. Thus, based on the data available at present, it is unclear whether sonorants can serve as targets of retroflex consonant harmony in other domains (e.g., non-initial /…C_1…C_2…/ sequences). This question must be left to future research. Cf. footnote 5 on page 310.

c. Retroflex harmony does not target (initial) sonorants[4]

naʈʰ-ɳaː	'to flee'	nuːɳ	'salt'
liʈ-ɳaː	'to lie down'	loɽ	'need, want'
roɖaː	'shaven head'	raːɽ	'quarrel'

The lack of interaction between obstruents and sonorants is predicted in part by the fixed ranking of CORR-T↔Ḍʰ over CORR-T↔Ṛʰ in (4). The former constraint demands correspondence between consonants that differ at most in laryngeal and coronal features. As a result, corresponding consonants must agree in manner/stricture features including [±son]. The latter constraint, which demands correspondence between all coronal consonants without regard for sonorancy, is ranked below the former constraint due to the subset-superset relation that holds between them.

The complete constraint ranking that predicts the attested similarity effect concerning obstruents and sonorants is presented in Tableaus (9) and (10). For the sake of simplifying the Tableaus and reducing clutter, I conflate the coronal features [±ant] and [±dist] into a single feature bundle [±ant, ±dist] in all of the constraints. For instance, the constraint IDENT-C_RC_L[−ant,−dist] represents both IDENT-C_RC_L[−ant] and IDENT-C_RC_L[−dist]. Output forms that would violate either one of the independent constraints incur a single violation of the conflated constraint; those that would violate both of the independent constraints incur a double violation of the conflated constraint. Here and elsewhere, subscript indices represent correspondence relations (e.g., [t_x … $ɖ_x$]), or the lack thereof (e.g., [t_x … $ɖ_y$]).

(9) Retroflex consonant harmony between plosives (e.g., Malto [ʈuːɖ] ‘tiger’)

/t … ɖ/	ID-IO [−ant, −dist]	CORR-T↔Ḍ^h	ID-C$_R$C$_L$ [−ant, −dist]	ID-IO [+ant, +dist]	CORR-T↔Ṛ^h
a. t$_x$ … ɖ$_y$		*!			*
b. t$_x$ … ɖ$_x$			*!*		
☞ c. ʈ$_x$ … ɖ$_x$				**	
d. ʈ$_x$ … ɖ$_y$		*!		**	*
e. t$_x$ … d$_x$	*!*				

(10) No harmony between sonorants and plosives (e.g., Malto [taɽe] ‘grinding stone’)

/t … ɽ/	ID-IO [−ant, −dist]	CORR-T↔Ḍ^h	ID-C$_R$C$_L$ [−ant, −dist]	ID-IO [+ant, +dist]	CORR-T↔Ṛ^h
☞ a. t$_x$ … ɽ$_y$					*
b. t$_x$ … ɽ$_x$			*!*		
c. ʈ$_x$ … ɽ$_x$				*!*	
d. ʈ$_x$ … ɽ$_y$				*!*	*
e. t$_x$ … r$_x$	*!*				

Tableau (9) demonstrates retroflex consonant harmony between plosives. Candidates (9)(a) and (d) are both eliminated because they fail to enforce correspondence between segments differing at most in laryngeal and coronal features, thereby violating CORR-T↔Ḍ^h. Candidate (9)(b) satisfies the demand for correspondence but violates high-ranked IDENT-CC[−ant, −dist], which demands agreement for retroflex features between correspondent consonants. Candidate (9)(e) avoids this violation by means of progressive de-retroflexion, but this incurs a fatal violation of IDENT-IO[−ant, −dist], which demands input-output faithfulness to retroflex features. (9)(c) emerges as the winning candidate because it satisfies the demand for correspondence and the demand for agreement in retroflexion, and because it does so only at the expense of low-ranked IDENT-IO[+ant,+dist], which demands faithfulness to non-retroflex features. The same results apply to other coronal plosive pairs, regardless of laryngeal distinctions (e.g., /t…ʈ/, /tʰ…ʈ/, /d…ɖʰ/, /d…ʈ/, etc.).

In Tableau (10) the same constraint ranking predicts that retroflex consonant harmony will not apply between plosives and sonorants. Correspondence between plosives and sonorants is enforced only by low-ranked CORR-T↔Ṛʰ, but not by high-ranked CORR-T↔Ḍʰ. Candidate (10)(b) satisfies the low-ranked demand for correspondence but fails to satisfy the high-ranked demand for retroflex agreement under correspondence. Candidate (10)(c) exhibits retroflex consonant harmony, thereby satisfying both the demand for correspondence and the demand for agreement under correspondence. Nevertheless, it is eliminated for violating input-output faithfulness, which takes precedence over correspondence between obstruents and sonorants. The violation of input-output faithfulness also proves fatal for (10)(d) and (e). This leaves the faithful candidate (10)(a) as the winner. The same results apply to other potential obstruent-sonorant and sonorant-obstruent pairs (e.g., /t…ɳ/, /d…ɭ/, /n…ɖ/, /l…ʈʰ/, etc.).

The proposed analysis accounts for the fact that retroflex consonant harmony holds between two obstruents but not between obstruents and sonorants. However, it also predicts that harmony should hold between two sonorants. Sequences such as /n…ɳ/, /l…ɭ/ and (possibly) /r…ɽ/ contain segments that differ at most in retroflexion. As a result they fall under the purview of the same constraints that drive harmony between plosives. With very few exceptions, however, retroflex sonorants do not occur morpheme-initially in South Asian languages. Thus, retroflex consonant harmony may be overridden by independent markedness constraints banning initial retroflex sonorants. For the purpose of the present analysis I assume a single markedness constraint, *RetroSon/#__, which prohibits retroflex sonorants in morpheme-initial position. When this constraint outranks the constraints that drive harmony, it has the effect of preventing retroflex harmony between sonorants, as shown in (11).[5]

(11) Contextual markedness overrides retroflex consonant harmony between sonorants (e.g., Panjabi [nuːɳ] 'salt').

/n … ɳ/	*RetroSon / #__	ID-IO [−ant, −dist]	CORR-T↔Ḍʰ	ID-C_RC_L [−ant, −dist]	ID-IO [+ant, +dist]
a. n_x … $ɳ_y$			*!		
☞ b. n_x … $ɳ_x$				**	
c. $ɳ_x$ … $ɳ_x$	*!				**
d. $ɳ_x$ … $ɳ_y$	*!		*		**
e. n_x … n_x		*!*			

Candidate (11)(a) is eliminated for violating correspondence between segments that agree in manner. Candidate (11)(c) satisfies the constraints demanding both correspondence and agreement. However, this candidate is eliminated by the undominated contextual markedness constraint banning initial retroflex sonorants, as is candidate (11)(d). Candidate (11)(e) is

---

[5] This analysis suggests that pairs of coronal sonorants might exhibit retroflex consonant harmony if/when they occur in non-initial positions (e.g., /...n…ɳ…/ → […ɳ…ɳ…]). A possible example of this type is Panjabi /nanaːɳ/ ~ /naɳaːɳ/ 'husband's sister' (Goswami, 2000, p. 202). Unfortunately, the data available at present is not sufficient to confirm whether examples of this type are systematic or exceptional, or whether they might be attributed to other factors independent of consonant harmony. This is partly because the present study has focused primarily on the co-occurrence of consonants in word-initial #C_1V(N)C_2 sequences. The fact that retroflex consonant harmony appears to be strictly root-internal in South Asia, combined with the fact that roots (and other morphemes) are seldom longer than one or two syllables, means that conclusive evidence bearing on this issue would be difficult to find even had the study examined co-occurrence patterns over longer domains.

eliminated for violating faithfulness to input retroflexion. Thus, the optimal candidate is (11)(b), which satisfies the demand for correspondence but sacrifices retroflex agreement in order to avoid a word-initial retroflex sonorant.[6]

Dardic languages such as Indus Kohistani and Kalasha exhibit more subtle similarity effects within the class of obstruents. In these languages, retroflexion is contrastive for plosives, affricates and fricatives. Both languages exhibit coronal consonant harmony between obstruents of the same manner (i.e., plosive-plosive, affricate-affricate and fricative-fricative pairs). In addition, Indus Kohistani exhibits coronal harmony between sibilants of any kind, including affricate-fricative and fricative-affricate pairs. However, neither language exhibits consonant harmony between plosives and sibilants (i.e., plosive-fricative, plosive-affricate, fricative-plosive or affricate-plosive pairs). Thus, by hypothesis, there is greater similarity between sibilant affricates and fricatives than between sibilants of any kind and non-sibilant plosives. Representative examples from Indus Kohistani are listed in (12).

(12) Retroflex consonant harmony in Indus Kohistani (data from Zoller, 2005)

a. Retroflex harmony between plosives

ʈʌʈúː	‘a small horse’	< *t…ʈ
ʈaɳɖáṽ	‘to beat’	< *t…ɖ

---

[6] If correspondence (CORR-T↔Ḍʰ) and agreement (ID-$C_R C_L$[−ant, −dist]) are unranked relative to one another, or if their relative ranking is inverted, then (11)(a) would emerge as the optimal candidate. Empirically, the result of these rankings is indistinguishable from that of the proposed ranking, since (11)(b) and (11)(a) differ only in the presence or absence of a hypothesized correspondence relation. Ambiguity of this type is a consequence of segregating correspondence from agreement and remains an outstanding issue for the ABC model. See Hansson (2010, p. 333ff.) concerning the segregation of correspondence and agreement constraints, and concerning the prospect of replacing both with a single family of ANTICIPATE[F] constraints.

b. Retroflex harmony between affricates

ʈʂìʈʂʰ	'grey, spotted'	< *ts/tʃ…ʈʂ
ʈʂʰiʈʂáṽ	'to learn'	< *ts/tʃ…ʈʂ

c. Retroflex harmony between fricatives

ʂìʂ	'a head'	< *s/ʃ…ʂ
ʂùːʂ	'decent, fine, proper'	< *s/ʃ…ʂ

d. Retroflex harmony between affricates and fricatives

ʈʂoːʂáṽ	'to wring out'	< *ts/tʃ…ʂ
ʐʌmʈʂúː	'a son-in-law'	< *z/ʒ…ʈʂ

e. No retroflex harmony between plosives and affricates

tsaʈáṽ	'to lick'
tʃʌʈúː	'a grater for spices'
taʈʂʰáṽ	'to carve'

f. No retroflex harmony between plosives and fricatives

siʈìː	'a whistle'
ʃòʈʰ	'a bump, swelling'
dùːʂ	'a sin'

The constraint ranking that accounts for agreement between obstruents with identical manners of articulation, like the Indus Kohistani forms in (12)(a–c), is essentially the same as that of previous tableaus. The constraint CORR-T↔Ḍʰ, which demands correspondence between segments that differ at most in laryngeal features and coronal place, outranks other correspondence constraints including CORR-T↔Ẓʰ, which demands correspondence between obstruents of any kind, regardless of manner. This is illustrated in Tableau (13).

(13) Retroflex consonant harmony between obstruents of the same manner (e.g., Indus Kohistani [ʂìʂ] 'head')

/ʃ … ʂ/	ID-IO [−ant, −dist]	CORR-T↔Ḍʰ	ID-C$_R$C$_L$ [−ant, −dist]	ID-IO [+ant, +dist]	CORR-T↔Ẓʰ
a. $ʃ_x$ … $ʂ_y$		*!			*
b. $ʃ_x$ … $ʂ_x$			*!		
☞ c. $ʂ_x$ … $ʂ_x$				*	
d. $ʂ_x$ … $ʂ_y$		*!		*	*
e. $ʃ_x$ … $ʃ_x$	*!				

In (13), the input consists of two obstruents with identical manners of articulation (in this case, fricatives) that disagree in retroflexion. Candidates (13)(a) and (d) are both eliminated for violating CORR-T↔Ḍʰ, which demands correspondence between consonants with the same manner. Candidate (13)(b) satisfies the demand for correspondence but fails to exhibit retroflex agreement, thereby incurring a fatal violation of ID-C$_R$C$_L$[−ant, −dist]. Candidate (13)(e) is eliminated for violating faithfulness to input retroflexion. This leaves (13)(c) as the winning candidate. It satisfies both the demand for correspondence and the demand for agreement under correspondence, and does so only at the expense of faithfulness to non-retroflex features. The same ranking also accounts for other same-manner pairs including plosive-plosive combinations, like those in (12)(a), and affricate-affricate combinations, like those in (12)(b).

The remaining similarity effects in Indus Kohistani can be accounted for if the constraint that demands correspondence between sibilants (CORR-TS↔Ẓʰ) is ranked above that which demands correspondence between non-continuants (CORR-T↔J̣ʰ). According to the similarity-based correspondence hierarchy in (4), both of these must be ranked below CORR-

T↔Ḍʰ and above CORR-T↔Ẓʰ. This yields the ranking: CORR-T↔Ḍʰ ⟩ CORR-TS↔Ẓʰ ⟩ CORR-T↔J̣ʰ ⟩ CORR-T↔Ẓʰ. Faithfulness to non-retroflex features must be ranked below CORR-TS↔Ẓʰ but above all lower-ranked correspondence constraints. IDENT-C_RC_L[−ant, −dist], which demands retroflex agreement between correspondent segments, must outrank faithfulness to non-retroflex features. IDENT-IO[−ant, −dist], which demands faithfulness to input retroflexion, must be undominated. The effects of this ranking are demonstrated in Tableaus (14) and (15).[7]

(14) Retroflex consonant harmony between affricates and fricatives (e.g., Indus Kohistani [ʈʂoːʂáṽ] 'to wring out')

/ʧ … ʂ/	ID-IO [−ant, −dist]	CORR-TS↔Ẓʰ	ID-C_RC_L [−ant, −dist]	ID-IO [+ant, +dist]	CORR-T↔Ẓʰ
a. $ʧ_x$ … $ʂ_y$		*!			*
b. $ʧ_x$ … $ʂ_x$			*!		
☞ c. $ʈʂ_x$ … $ʂ_x$				*	
d. $ʈʂ_x$ … $ʂ_y$		*!		*	*
e. $ʧ_x$ … $ʃ_x$	*!				

[7] To avoid clutter, CORR-T↔Ḍʰ and CORR-T↔J̣ʰ are omitted from Tableaus (14) and (15). CORR-T↔Ḍʰ plays no role in these examples because co-occurring input obstruents do not share the same manner of articulation. Omitting CORR-T↔J̣ʰ does not affect the outcome either, because it is ranked immediately above CORR-T↔Ẓʰ, and any form that violates the former also violates the latter.

(15) No retroflex consonant harmony between plosives and sibilants (e.g., Indus Kohistani [taʈʂ^{h}áṽ] ‘to carve’)

/t … ʈʂ/	ID-IO [−ant, −dist]	CORR-TS↔Ẓ^{h}	ID-C_RC_L [−ant, −dist]	ID-IO [+ant, +dist]	CORR-T↔Ẓ^{h}
☞ a. t_x … ʈʂ$_y$					*
b. t_x … ʈʂ$_x$			*!*		
c. ʈ$_x$ … ʈʂ$_x$				*!*	
d. ʈ$_x$ … ʈʂ$_y$				*!*	*
e. t_x … ʦ$_x$	*!*				

Tableau (14) demonstrates retroflex consonant harmony between affricates and fricatives. Candidates (14)(a) and (14)(d) are eliminated for failing to satisfy CORR-TS↔Ẓ^{h}, which demands correspondence between segments that agree with respect to [±strident]. Candidate (14)(b) satisfies the demand for correspondence but violates high-ranked IDENT-C_RC_L[−ant, −dist], which enforces agreement for retroflexion between correspondent segments. Candidate (14)(e) avoids this violation by means of de-retroflexion, but it incurs a fatal violation of faithfulness to input retroflexion. (14)(c) emerges as the optimal candidate because it satisfies agreement under correspondence only at the expense of low-ranked faithfulness to non-retroflex features. The same results apply to other affricate-fricative and fricative-affricate combinations (e.g., /ʦ…ʂ/, /ʦ^{h}…ʂ/, /ʧ…ʐ/, /s…ʈʂ/, /ʃ…ʈʂ/, etc.).

In Tableau (15), the same ranking predicts that retroflex consonant harmony will not apply between plosives and sibilants. Notice that CORR-TS↔Ẓ^{h} does not demand correspondence between plosives and affricates or plosives and fricatives because it only applies to pairs of segments that agree with respect to [±strid]. In Tableau (15), the only constraint demanding correspondence between sibilants and non-sibilants is low-ranked CORR-T↔Ẓʰ. Candidate (15)(b) satisfies this constraint but fails to satisfy dominant IDENT-$C_RC_L[-\text{ant}, -\text{dist}]$, which demands agreement of retroflex features under correspondence. Candidate (15)(c) satisfies both the demand for correspondence and the demand for agreement under correspondence. Nevertheless, it is eliminated for violating faithfulness to non-retroflex features, which takes precedence over the demand for correspondence between sibilants and non-sibilants. Faithfulness to non-retroflex features also eliminates candidate (15)(d) and faithfulness to input retroflexion eliminates (15)(e). As a result, the fully faithful candidate in (15)(a) emerges as the optimal candidate. The same results apply to other plosive-sibilant and sibilant-plosive combinations (e.g., /t…ʂ/, /d…ʈʂ/, /s…ʈ/, /ʧ…ɖ/, etc.).

In addition to retroflex consonant harmony, co-occurring sibilants in Indus Kohistani (and Kalasha) show a strong tendency toward palatal and dental agreement. Representative examples from Indus Kohistani are listed in (16).[8]

---

[8] There are some exceptions to agreement between dental and palatal sibilants in Indus Kohistani and Kalasha (e.g., IK /sʌ̀ʧ/ 'true'; /ʃaːzaːdà/ 'prince'; Kalasha /ʧitse maik/ 'to talk about this and that'; /suʃ/ 'needle'). If the exceptions are not principled, then the pattern of agreement could be described as gradient. The same is true of some cases of retroflex consonant harmony reviewed in Chapter 3. Gradient co-occurrence restrictions might be best modeled within a Stochastic OT framework (e.g., Boersma & Hayes, 2001). The present analysis is concerned primarily with deriving similarity effects and other typological properties of coronal harmony systems in South Asia. Thus, the issue of gradience is not explored here. See Martin (2005) for an ABC account of gradient co-occurrence restrictions in Navajo within a Stochastic OT framework.

(16) Palatal harmony between sibilants in Indus Kohistani (data from Zoller, 2005)

a. Palatal agreement between sibilants

ʧiːʧʎkʰ	‘smallpox’	*ʦ…ʧ, *ʧ…ʦ, etc.
ʧuʃtìː	‘absorption’	*ʦ…ʃ, *ʃ…ʦ, etc.
ʃã̀ːʧ	‘a kind of mungo’	*s…ʧ, *ʧ…s, etc.
ʃòːʃaː	‘decoration’	*s…ʃ, *ʃ…s, etc.

b. Dental agreement between sibilants[9]

ʦíʦʰⁱ	‘nipple, breast’
ʦã̀z	‘soft’
zʰã́ːz	‘a branch of a holm oak’

In the case of retroflex consonant harmony, there is abundant historical-comparative evidence to support a pattern of regressive assimilation in which retroflex segments dominate palatals and dentals. Unfortunately, there is little evidence to shed light on matters of directionality and dominance governing the co-occurrence of palatal and dental sibilants (see discussion in §3.3.1.2 and §3.6.3.2). Assuming that the co-occurrence of dental and palatal sibilants is avoided through regressive palatal assimilation, along the same lines as retroflex harmony, then the co-occurrence pattern in (16) is a natural consequence of the foregoing

[9] The present study found no evidence of dental consonants triggering dental harmony in any South Asian language. In §3.6.3.2 it was suggested that dental-dental configurations are not the product of assimilation. Rather, they are the configurations that remain unaffected by retroflex and palatal harmony. If so, then they only arise in output forms through faithfulness to input forms. Like all other input forms that show agreement for coronal features (i.e., palatal-palatal and retroflex-retroflex inputs), dental-dental configurations are expected to surface faithfully (all things being equal). This is because they will never violate IDENT-CC constraints pertaining to coronal features. As a result, any unfaithful mappings will always incur unwarranted violations of input-output faithfulness.

analysis. In addition to retroflex consonant harmony, the analysis sketched above also predicts regressive palatal harmony targeting dental sibilants. This is illustrated in (17).

(17) Palatal consonant harmony between sibilants (e.g., Indus Kohistani [ʃã̀ːtʃ] ‘a kind of mungo’)

/s … tʃ/	ID-IO [−ant, −dist]	CORR-TS↔Ẓʰ	ID-C_RC_L [−ant, −dist]	ID-IO [+ant, +dist]	CORR-T↔Ẓʰ
a. s_x … $tʃ_y$		*!			*
b. s_x … $tʃ_x$			*!		
☞ c. $ʃ_x$ … $tʃ_x$				*	
d. $ʃ_x$ … $tʃ_y$		*!		*	*
e. s_x … ts_x	*!				

The constraints and constraint rankings in (17) are exactly the same as those in (14) and (15), which account for retroflex consonant harmony between sibilants in Indus Kohistani. Given an input consisting of a dental sibilant followed by a palatal sibilant, the same constraint ranking predicts regressive palatal assimilation in (17). This is because the constraint that drives retroflex agreement (i.e., ID-C_RC_L[−ant, −dist]) demands agreement, not only for [−dist], but also for [−ant], which is the feature that distinguishes palatals from dentals (cf. Table 32, p. 299).[10] In (17), the input sibilants disagree with respect to coronal place and

---

[10] The prediction that retroflex harmony entails palatal harmony follows from the phonological features assumed in the present analysis, not from anything inherent in the ABC model. The prediction is a consequence of treating retroflex and palatal segments as a natural class defined by the feature [−ant]. This prediction seems desirable (or at least unproblematic) in the case of Dardic languages, such as Indus Kohistani and Kalasha, which are the focus of the current analysis. If necessary, it could be avoided by assuming a feature system in which retroflex and palatal segments do not constitute a natural class. For instance, agreement for a privative feature such as [retroflex] would not entail palatal agreement. Similarly, the prediction might be avoided within the current feature system if

continuancy, but agree with respect to all other features including stridency. Thus, they fall under the purview of CORR-TS↔Ẓʰ, which demands correspondence between segments that disagree *at most* in coronal place, laryngeal features and continuancy. Candidates (17)(a) and (d) both incur fatal violations of this constraint. Candidate (17)(c), satisfies correspondence but fails to enforce agreement for [−ant], thereby incurring a fatal violation of ID-C_RC_L[−ant,−dist]. Candidate (17)(e) avoids this violation through de-palatalization, but is eliminated for violating faithfulness to input specifications of [−ant]. The winning candidate is (17)(c), which satisfies both correspondence and agreement for [−ant] through regressive palatal assimilation. The same results apply to other sibilant-sibilant pairs in which dental sibilants are followed by palatal sibilants (e.g., /ʦ…ʧ/, /ʦ…ʃ/, /s…ʃ/, etc.). Other potential inputs, including those consisting of initial palatal sibilants followed by dental sibilants (e.g., /ʧ…ʦ/, /ʃ…s/, etc.) and initial retroflex sibilants followed by palatals (e.g., /ʈʂ…ʃ/, /ʂ…ʃ/, etc.), are discussed in §5.4.1, below.

In summary, the similarity effects observed in retroflex consonant harmony systems in South Asia can be captured through the interaction of CORR-C↔C, IDENT-CC, and IDENT-IO constraints within the ABC model of long-distance consonant agreement. The following section briefly discusses a few outstanding issues.

---

retroflex agreement is modelled only as agreement for [−dist]. In this case, we would have to assume that changes in anteriority (i.e., when dentals assimilate to retroflexes) are not driven by agreement for [−ant]. They could be regarded as secondary changes required to satisfy agreement for [−dist], given that all [−dist] segments are also [−ant] within the phonological systems of the languages under discussion.

5.4 Outstanding issues

The analysis sketched in the preceding section demonstrates that the ABC model is able to provide a straightforward account of similarity effects in the retroflex consonant harmony systems of South Asian languages. A few outstanding issues deserve mention. The following sub-sections briefly discuss issues concerning dominance and directionality (§5.4.1), the role of similarity in conditioning the co-occurrence of Kalasha sibilants (§5.4.2) and the range of possible similarity effects predicted by the ABC model (§5.4.3).

5.4.1 Dominance and directionality

Retroflex consonant harmony clearly involves regressive assimilation in South Asia. However, the status of progressive assimilation remains speculative. Languages with retroflex harmony generally lack both T-Ṭ and Ṭ-T sequences. As discussed in §3.6.3.2, the avoidance of Ṭ-T configurations could be attributed to progressive retroflex assimilation (Ṭ-T → Ṭ-Ṭ) or regressive de-retroflexion (Ṭ-T → T-T). However, as we have seen, there is no historical-comparative evidence for either of these processes (at least not in the case of root-internal, similarity-sensitive harmonies). For the sake of discussion, let us assume that retroflex consonant harmony in South Asia is like palatal harmony in the Dravidian languages, Pengo and Kuvi (§3.1.2.2). That is, let us assume that it is strictly regressive and that retroflexes consistently dominate non-retroflex coronals. This interpretation predicts that hypothetical Ṭ-T sequences would be preserved intact if they were to arise. The strictly regressive nature of harmony would prevent progressive retroflexion (Ṭ-T ↛ Ṭ-Ṭ) and the recessive nature of non-retroflex coronals would prevent regressive de-retroflexion (Ṭ-T ↛ T-T). The ABC account sketched in the preceding section predicts these effects by encoding regressive directionality on the IDENT-CC constraints that drive long-distance agreement. This is illustrated in (18).

(18) No progressive retroflex consonant harmony with directional IDENT-CC constraints

/ʈ … d/	ID-IO [−ant, −dist]	CORR-T↔Ḍ^h	ID-C$_R$C$_L$ [−ant, −dist]	ID-IO [+ant, +dist]	CORR-T↔Ṛ^h
a. ʈ$_x$ … d$_y$		*!			*
☞ b. ʈ$_x$ … d$_x$					
c. ʈ$_x$ … ɖ$_x$				*!*	
d. ʈ$_x$ … ɖ$_y$		*!		**	*
e. t$_x$ … d$_x$	*!*				

The constraints and their rankings in (18) are the same as those of Tableau (9) in the preceding section, which accounted for regressive retroflex assimilation between plosives (i.e., T-Ṭ → Ṭ-Ṭ). This analysis predicts that input configurations of the type Ṭ-T, with initial retroflex plosives, will surface faithfully. In Tableau (18), candidates (a) and (d) are both eliminated for failing to enforce correspondence between obstruents with the same manner. All other candidates satisfy the demand for correspondence. Candidate (18)(b) shows disagreement for retroflexion while candidate (18)(c) shows agreement. However, the constraint ID-C$_R$C$_L$[−ant, −dist] only evaluates agreement from right to left. (18)(b) does not violate this constraint, despite its lack of retroflex agreement, because the rightmost correspondent is not retroflex (i.e., it does not bear the features [−ant] or [−dist]). As a result, the candidate with retroflex agreement in (18)(c) has no advantage over (18)(b). It is eliminated because its violation of faithfulness to input features is unwarranted. Candidate (18)(e) is also eliminated for violating faithfulness to input features. Thus, candidate (18)(b), which exhibits correspondence without progressive retroflex assimilation, emerges as the optimal candidate.

The same results hold for coronal sibilant harmony in Dardic languages such as Indus Kohistani. The constraint ranking proposed for Indus Kohistani in Tableau (17) of the preceding section predicts that an input containing a palatal sibilant followed by a dental sibilant (e.g., /ʃ…s/) will emerge faithfully without progressive palatal assimilation or regressive de-palatalization, as shown in (19).

(19) No progressive palatal consonant harmony with directional IDENT-CC constraints

/ʃ … s/	ID-IO [−ant, −dist]	CORR-TS↔Ẓʰ	ID-C_RC_L [−ant, −dist]	ID-IO [+ant, +dist]	CORR-T↔Ẓʰ
a. $ʃ_x$ … s_y		*!			*
☞ b. $ʃ_x$ … s_x					
c. $ʃ_x$ … $ʃ_x$				*!	
d. $ʃ_x$ … $ʃ_y$		*!		*	*
e. s_x … s_x	*!				

Once again, candidates (19)(a) and (d) are eliminated for violating the demand for correspondence (CORR-TS↔Ẓʰ). Candidates (19)(b) and (c) satisfy correspondence but ID-C_RC_L[−ant, −dist] does not demand agreement between correspondent segments in this case because the rightmost sibilant is neither [−ant] nor [−dist]. As a result, candidate (19)(c), which exhibits progressive palatal assimilation, is eliminated for an unwarranted violation of faithfulness to input features. Candidate (19)(e), which exhibits regressive de-palatalization, also incurs a fatal violation of input-output faithfulness. Thus, the optimal output is the faithful candidate in (19)(b), which exhibits correspondence without agreement.

A similar outcome is predicted for input configurations of the type Ṣ-Š, which contain initial retroflex sibilants followed by palatal sibilants. In theory, input Ṣ-Š configurations might be subject to progressive retroflex assimilation (e.g., Ṣ-Š → Ṣ-Ṣ) or regressive palatal assimilation (e.g., Ṣ-Š → Š-Š). Under the current analysis neither of these outcomes is expected. Rather, input Ṣ-Š configurations are expected to surface faithfully, much like Ṭ-T and Š-S configurations in the preceding tableaus. This is illustrated in (20).

(20) No assimilation in Ṣ-Š configurations with directional IDENT-CC constraints

/ʂ … ʃ/	ID-IO [−ant, −dist]	CORR-TS↔Ẓ^h	ID-C_RC_L [−ant, −dist]	ID-IO [+ant, +dist]	CORR-T↔Ẓ^h
a. ʂ$_x$ … ʃ$_y$		*!			*
☞ b. ʂ$_x$ … ʃ$_x$					
c. ʃ$_x$ … ʃ$_x$	*!				
d. ʃ$_x$ … ʃ$_y$	*!	*			*
e. ʂ$_x$ … ʂ$_x$				*!	

Candidates (20)(b), (c) and (e) all satisfy the demand for correspondence between sibilants (CORR-TS↔Ẓ^h). Recall that ID-C_RC_L[−ant, −dist] can be decomposed into ID-C_RC_L[−ant] and ID-C_RC_L[−dist]. The co-occurring sibilants in each candidate agree with respect to [−ant]. Thus, no candidate that satisfies correspondence violates ID-C_RC_L[−ant]. In candidates (b) and (c), the rightmost sibilant is [+dist]. Thus, ID-C_RC_L[−dist] does not apply to these candidates despite the fact that they exhibit correspondence. Without the pressure for further agreement (beyond agreement for [−ant]), violations to input-output faithfulness in candidates (c) and (e) are fatal. As a result, candidate (b), which satisfies correspondence while remaining faithful to input features, emerges as the optimal candidate.

The prediction that retroflex consonant harmony in South Asia and palatal sibilant harmony in Dardic are strictly regressive is consistent with the system of palatal harmony in

Pengo and Kuvi, where disharmonic T-Č configurations are subject to regressive assimilation (i.e., T-Č → Č-Č) but disharmonic Č-T configurations are preserved intact (§3.1.2.2). At the very least, therefore, systems of this kind must be possible. However, unlike palatal harmony in Pengo and Kuvi, most of the consonant harmony systems examined in the current study do not exhibit disharmonic surface forms with initial retroflex or palatal segments (i.e., *Ṭ-T, *Ṣ-S, *Ṣ-Š, *Š-S, etc.). The avoidance of these configurations may not be an accident of history (§3.6.3.2). If it turns out that these hypothetical input configurations are subject to repair via progressive assimilation, then this effect can be accommodated within the ABC model by including 'progressive' agreement constraints (e.g., ID-C_LC_R[−ant,−dist]) alongside 'regressive' agreement constraints (e.g., ID-C_RC_L[−ant,−dist]), or by collapsing progressive and regressive agreement constraints into a single non-directional (=bi-directional) constraint (e.g., ID-CC[−ant, −dist]). The latter option is illustrated in (21) in relation to a hypothetical input with an initial retroflex plosive.

(21) Progressive retroflex consonant harmony with non-directional IDENT-CC constraints

/ʈ … d/	ID-IO [−ant,−dist]	CORR-T↔Ḍ^h	ID-CC [−ant,−dist]	ID-IO [+ant,+dist]	CORR-T↔Ṛ^h
a. ʈ$_x$ … d$_y$		*!			*
b. ʈ$_x$ … d$_x$			*!*		
☞ c. ʈ$_x$ … ɖ$_x$				**	
d. ʈ$_x$ … ɖ$_y$		*!		**	*
e. t$_x$ … d$_x$	*!*				

Tableau (21) is exactly like (18) except that the constraint that demands agreement for retroflexion between correspondent segments (i.e., ID-CC[−ant,−dist]) now evaluates agreement from right to left *and from left to right*. As a result, candidate (21)(b) (which corresponds to the winning candidate in (18)) is eliminated for exhibiting correspondence without retroflex agreement. Its closest competitor, candidate (21)(c), emerges as the optimal candidate by satisfying the demand for agreement under correspondence via progressive retroflex assimilation. Assuming a non-directional formulation of IDENT-CC constraints, progressive retroflex assimilation would also apply to inputs containing retroflex sibilants followed by non-retroflex sibilants (e.g., /ʂ…s/, /ʈʂ…ʦ/, /ʂ…ʃ/, /ʈʂ…ʧ/, etc.), while progressive palatal assimilation would apply to inputs containing palatal sibilants followed by dental sibilants (e.g., /ʃ…s/, /ʧ…ʦ/, etc.).[11]

By encoding directionality on the IDENT-CC constraints that drive harmony, Rose & Walker's (2004) ABC model is able to account for either directional or bi-directional assimilation. However, regardless of whether retroflex consonant harmony is regressive or bi-directional, the possibility of progressive de-retroflexion is ruled out only by faithfulness to retroflexion, not by directionality. This can be seen in Tableau (9), repeated here as (22), which assumes a strictly regressive pattern of retroflex assimilation.

---

[11] Under the proposed analysis, regressive palatal assimilation is not expected for input retroflex-palatal configurations (e.g., Ṣ-Š, Ç-Č, etc.) regardless of whether we assume a directional or non-directional formulation of IDENT-CC constraints. This is because changes of the type Ṣ → Š violate IDENT-IO[−dist], which is undominated, while changes of the type Š → Ṣ violate IDENT-IO[+dist], which is ranked below the former constraint. This may or may not be a desirable prediction. The current study found no evidence of how hypothetical inputs of these types are actually treated.

(22) Regressive retroflex consonant harmony between plosives (repeated from (9))

/t … ɖ/	ID-IO [−ant, −dist]	CORR-T↔Ḍʰ	ID-C$_R$C$_L$ [−ant, −dist]	ID-IO [+ant, +dist]	CORR-T↔Ṛʰ
a. t_x … $ɖ_y$		*!			*
b. t_x … $ɖ_x$			*!*		
☞ c. $ʈ_x$ … $ɖ_x$				**	
d. $ʈ_x$ … $ɖ_y$		*!		**	*
e. t_x … d_x	*!*				

Candidates (22)(c) and (22)(e) both satisfy the demand for correspondence and the demand for feature agreement under correspondence. (22)(c) satisfies the demand for agreement by means of regressive retroflex assimilation, whereas (22)(e) does so by means of progressive de-retroflexion. (22)(c) emerges as the winner, not because it applies regressive assimilation, but because it satisfies agreement at the expense of low-ranked faithfulness to non-retroflex features (i.e., ID-IO[+ant, +dist]), whereas (22)(e) does so at the expense of high-ranked faithfulness to retroflex features (i.e., ID-IO[−ant, −dist]). Faithfulness to retroflex features is valued over faithfulness to non-retroflex features. In effect, progressive de-retroflexion is ruled out by a dominance effect, not by strict regressive assimilation. This is an unsatisfactory result if the intent is to motivate regressive assimilation independent of dominance and stem control.

In response to this problem, Hansson (2010) has proposed an alternative account of directionality within the ABC model. In Hansson's account, directionality is encoded in the correspondence relation itself. Rose & Walker (2004) assume that correspondence relations are symmetrical. When consonants stand in a correspondence relation they are correspondents of *one another* ($C_1 \leftrightarrow C_2$). Hansson's formulation assumes that correspondence relations are asymmetrical. In a sequence of the type $C_1 \ldots C_2$, correspondence constraints will enforce a correspondence mapping from C_2 to C_1 but not *vice versa* ($C_1 \leftarrow C_2$). This is intended to capture the generalization that correspondence relations are primarily anticipatory and that, all things being equal, consonant harmony is primarily regressive. However, it turns out that asymmetric correspondence alone is not sufficient to guarantee strict directionality. Hansson's account also requires the reformulation of IDENT-CC constraints as *targeted* constraints (cf. Wilson, 2000; 2001).

A full account of Hansson's (2010) approach to directionality will not be presented here, as it would require a lengthy digression from the current study's focus on similarity effects. Suffice it to say that the status of targeted constraints, as employed by Hansson, remains controvertial in OT (McCarthy, 2002). The best means of deriving strict directional effects remains an outstanding issue for the ABC model. This issue has little bearing on the current analysis where there is some evidence for both regressive directionality and dominance effects. However, the issue takes on greater significance for those languages that exhibit directional consonant harmony without dominance effects (e.g., Chumash sibilant harmony). A solution to this issue is beyond the scope of the present study.

5.4.2 Kalasha sibilants

Another outstanding issue is the degree to which similarity effects are responsible for the co-occurrence of sibilants in Kalasha. In Kalasha, affricate-affricate pairs and fricative-fricative pairs are clearly subject to retroflex consonant harmony (e.g., Kalasha /ɖʐ̺aʈʂ̺/ 'spirit beings'; /ʂ̺uʂ̺-ik/ 'to dry'), but not all affricate-fricative or fricative-affricate pairs show retroflex agreement. Thus, unlike Indus Kohistani, retroflex consonant harmony in Kalasha might be sensitive to the distinction between affricates and fricatives (e.g., Kalasha /ʧuʂ̺ik/ 'to suck', cf. Indus Kohistani /ʈʂ̺oːʂ̺áṽ/ 'to suck (out)', both of which are cognate with OIA /ʧuːʂ̺ati/ 'sucks'). If so, then this similarity effect can be accounted for within the ABC model by ranking constraints in such a way that harmony is enforced only between obstruents that share all manner/stricture features, including [±continuant]. The relevant constraint ranking and its effects are illustrated in (23) and (24). The ranking here is essentially the same as that for Indus Kohistani in (14) and (15) except that CORR-TS↔Ẓʰ, which demands correspondence between strident obstruents differing in continuancy (i.e., affricates and fricatives), is demoted to a position below faithfulness to all input coronal features, including non-retroflex features (i.e., ID-IO[+ant, +dist]).

(23) Retroflex consonant harmony between sibilants that agree in stricture (e.g., Kalasha [ʂ̺uʂ̺ik] 'to dry')

/ʃ … ʂ̺/	ID-IO [−ant, −dist]	CORR-T↔Ḍʰ	ID-C_RC_L [−ant, −dist]	ID-IO [+ant, +dist]	CORR-TS↔Ẓʰ
a. $ʃ_x$ … $ʂ̺_y$		*!			*
b. $ʃ_x$ … $ʂ̺_x$			*!		
☞ c. $ʂ̺_x$ … $ʂ̺_x$				*	
d. $ʂ̺_x$ … $ʂ̺_y$		*!		*	*
e. $ʃ_x$ … $ʃ_x$	*!				

(24) No retroflex consonant harmony between sibilants that disagree in stricture (e.g., Kalasha [tʃuʂik] ‘to suck’)

/tʃ … ʂ/	ID-IO [−ant, −dist]	CORR-T⟺Ḍʰ	ID-C_RC_L [−ant, −dist]	ID-IO [+ant, +dist]	CORR-TS⟺Ẓʰ
☞ a. $tʃ_x$ … $ʂ_y$					*
b. $tʃ_x$ … $ʂ_x$			*!		
c. $tʂ_x$ … $ʂ_x$				*!	
d. $tʂ_x$ … $ʂ_y$				*!	*
e. $tʃ_x$ … $ʃ_x$	*!				

The constraint ranking in (23) predicts that sibilants of the same manner will be subject to retroflex agreement in Kalasha. CORR-T⟺Ḍʰ demands correspondence between obstruents with identical manners of articulation (i.e., those differing at most in laryngeal and coronal place features). Candidates (23)(a) and (d) are both eliminated for violating this constraint. Candidate (23)(b) satisfies the demand for correspondence but violates the demand for retroflex agreement between correspondent segments (i.e., IDENT-CC[−ant, −dist]). Candidate (23)(e) avoids this violation by means of progressive de-retroflexion, but incurs a fatal violation of faithfulness to retroflex features (i.e., IDENT-IO[−ant, −dist]). (23)(c) emerges as the winning candidate because it satisfies the demand for correspondence and the demand for retroflex agreement, and because it does so only at the expense of faithfulness to non-retroflex features (i.e., IDENT-IO[+ant, +dist]).

In (24), the same constraint ranking predicts that sibilants with different manners will not be subject to retroflex agreement. The pressure for correspondence is weaker in these pairs. A failure to establish correspondence between affricates and fricatives violates low-ranked

CORR-TS⇔Ẓʰ, but not high-ranked CORR-T⇔Ḍʰ. Candidates (24)(c–e) are all eliminated for violating faithfulness to input coronal features, a change that is unwarranted without the pressure for correspondence and agreement. Candidate (24)(b) is faithful to input features, but exhibits correspondence without agreement. The winning candidate is the faithful candidate in (24)(a), which lacks correspondence, thereby avoiding the demand for agreement altogether.

The constraint ranking in (23) and (24) predicts that sibilants of the same manner will be subject to retroflex agreement, while sibilants that disagree in manner will not. However, as noted earlier (§4.3.1.2), the co-occurrence of affricates and fricatives in Kalasha exhibits an asymmetry that may not be reducible to similarity effects alone. If all affricate-fricative and fricative-affricate configurations are exempt from harmony, then (all things being equal) we would expect all of them to be well-attested. As it is, configurations involving palatal affricates with retroflex fricatives are well-attested (e.g., /ʧ…ʂ̻/, /ʂ̻…ʧ/, etc.), but configurations involving retroflex affricates with palatal fricatives are not (e.g., */ʈʂ̻…ʃ/, */ʃ…ʈʂ̻/, etc.). If this asymmetry is the result of a similarity effect, then it implies that retroflex affricates are somehow more similar to fricatives than palatal affricates are. If this turns out to be true, then the pattern could be modelled within the ABC framework, given an appropriate set of features to capture the relevant natural classes. However, the notion that retroflex affricates bear a stronger similarity than palatal affricates to the entire class of coronal fricatives seems doubtful, and there is no way to capture the relevant natural classes using the features assumed here (or using any other set of features that I am aware of).[12] Other factors apart from similarity may be involved,

[12] On the assumption that the asymmetry in Kalasha is a similarity effect respecting natural classes, the relevant classes would be as follows. On the one hand, there would be a natural class of affricates that includes palatal

though their precise nature remains unclear at present. A full account of the co-occurrence of Kalasha sibilants remains an outstanding issue for future research.[13]

5.4.3 Unattested similarity effects

Another outstanding issue concerns the range of possible similarity effects. The ABC model performs well in capturing the attested similarity effects in retroflex consonant harmony. Nevertheless, it may over-generate by predicting some (as yet) unattested similarity effects. Two predictions deserve mention here. The first concerns the interaction of manner features, specifically [±strid] and [±cont]; the second concerns the role of laryngeal features in coronal harmony systems. Each of these is discussed in turn below.

To begin with, the model of phonological features adopted here follows Clements (1999) in representing affricates as non-contoured strident stops. This representation predicts that affricates are as similar to plosives, with which they share the feature [−cont], as they are

---

affricates because all affricate-affricate pairs interact in harmony. This set would consist of {dental affricates, palatal affricates, retroflex affricates}. On the other hand, there would be a broader class of sibilants that excludes palatal affricates. This set would consist of {dental affricates, retroflex affricates, dental fricatives, palatal fricatives, retroflex fricatives} to the exclusion of {palatal affricates}. The members of the broader set would be deemed sufficiently similar to interact with one another in retroflex consonant harmony (thereby predicting the avoidance of */ʃ…ʈʂ/, */ʈʂ…ʃ/, etc.). Palatal affricates would be deemed sufficiently dissimilar from the members of that set, so as to avoid interaction with them (thereby predicting the preservation of /ʧ…ʂ/, /ʂ…ʧ/, etc.).

[13] Retroflex consonant harmony in Komi-Permyak may be sensitive to the affricate vs. fricative distinction (Kochetov, 2007). Thus, whatever their status in Kalasha, the similarity effect and associated constraint ranking discussed here may be independently attested elsewhere. However, it is worth noting that, like Kalasha, Komi-Permyak shows a preference for Č-Ṣ and Ṣ-Č configurations (with palatal affricates and retroflex fricatives) over *Ç-Š and *Š-Ç configurations (with retroflex affricates and palatal fricatives).

to fricatives, with which they share the feature [+strid]. The similarity-based correspondence hierarchy is expected to include both CORR-TS↔Ẓ^{h}, which demands correspondence between affricates and fricatives (i.e., segments that agree in [±strid] but disagree in [±cont]), and CORR-T↔J̣^{h}, which demands correspondence between affricates and plosives (i.e., segments that agree in [±cont] but disagree in [±strid]). These two constraints have no inherent ranking relative to one another because the class of segments specified in one does not constitute a subset of the other. Wherever the ranking of CORR-C↔C constraints cannot be determined on the basis of subset-superset relations, Rose & Walker (2004) assume that it may be subject to cross-linguistic variation.

In the case of Dardic languages like Indus Kohistani it is clear that CORR-TS↔Ẓ^{h} must outrank CORR-T↔J̣^{h}, since consonant harmony holds between affricates and fricatives, but not between affricates and plosives. However, the prospect of cross-linguistic variation predicts the possible occurrence of a language similar to Indus Kohistani with just the opposite ranking (i.e., CORR-T↔J̣^{h} ⟫ CORR-TS↔Ẓ^{h}). Such a language would have coronal place contrasts in plosives, affricates and fricatives, and would exhibit coronal harmony between plosives and affricates but not between affricates and fricatives. To the best of my knowledge, no such language exists.[14] The absence of this particular similarity effect might be an accidental gap. Alternatively, if stridency is inherently more salient than continuancy, then the ranking of

---

[14] Palatal harmony in Pengo and Kuvi might come close. In these languages palatal affricates trigger harmony in dental plosives, but not in dental fricatives. However, it is not clear that affrication is contrastive in these languages. The palatal affricates might be phonologically equivalent to plosives (with redundant affrication). Moreover, the non-participation of fricatives might simply reflect that fact that palatal fricatives do not occur in these languages. Thus, it is not clear that an analysis of palatal harmony in Pengo and Kuvi would require reference to the constraint CORR-T↔J̣^{h} at all, let alone the ranking CORR-T↔J̣^{h} ⟫ CORR-TS↔Ẓ^{h}.

CORR-TS↔Ẓʰ over CORR-T↔J̣ʰ in Indus Kohistani might reflect a universal trend. This remains an outstanding issue for further research.[15]

A more serious problem for similarity effects within the ABC model concerns the role of laryngeal features in coronal consonant harmony systems. In the ABC analysis sketched in §5.3, the highest ranked correspondence constraint was CORR-T↔Ḍʰ, which demands correspondence between segments that differ at most in retroflexion and laryngeal features. However, the model also predicts the occurrence of other constraints including: (i) CORR-T↔Ḍ, which demands correspondence between segments that disagree in voicing but not in aspiration; (ii) CORR-T↔Ṭʰ, which demands correspondence between segments that disagree in aspiration but not in voicing; and (iii) CORR-T↔Ṭ, which demands correspondence between segments that agree in all laryngeal features. Subset-superset relations demand that all of these constraints outrank CORR-T↔Ḍʰ, as shown in (25).

(25) Similarity-based correspondence hierarchy for laryngeal features

CORR-T↔Ṭ	⟫	CORR-T↔Ṭʰ	,	CORR-T↔Ḍ	⟫	CORR-T↔Ḍʰ
'same laryngeal'		'same voicing'		'same aspiration'		'any laryngeal'

The hierarchy in (25) makes at least one desirable prediction. Any language that enforces retroflex agreement between segments that differ in laryngeal features (i.e., low-

[15] It should be pointed out that issues of this kind are primarily issues concerning phonological feature theory, not issues concerning the ABC model *per se*. If it turns out that stops and affricates do not pattern as a class to the exclusion of fricatives, then the prediction that they might do so could be avoided simply by assuming a different set of features (i.e., one that avoids the unattested natural class) or by assuming that the feature distinguishing affricates and fricatives is only applicable to sibilants and does not extend to plosives. Cf. footnote 2 on page 296.

ranked CORR-T↔Ḍ^{h}) will also enforce agreement between segments that agree in laryngeal features (i.e., higher-ranked CORR-T↔Ḍ, CORR-T↔Ṭ^{h} and CORR-T↔Ṭ). However, if faithfulness to input features is ranked above CORR-T↔Ḍ^{h}, but below CORR-T↔Ṭ, then the model predicts that retroflex harmony (or other coronal harmonies) will exhibit sensitivity to laryngeal features. For instance, we might expect to find a language in which retroflex harmony applies only to consonant pairs that agree in laryngeal features. In such a system, harmony would apply to sequences such as /t…ʈ/, /d…ɖ/ or /t^{h}…ʈ^{h}/, but not to /t…ɖ/, /t…ʈ^{h}/, /t…ɖ^{h}/, or any other pair that exhibits a laryngeal mismatch. To the best of my knowledge, no system of this kind exists. The survey of retroflex consonant harmony in South Asia found that most systems are sensitive to similarity in terms of manner/stricture features, but not in terms of laryngeal features. This generalization appears to hold for minor place harmony systems cross-linguistically, whether coronal or dorsal (Rose & Walker, 2004, p. 485). Thus, the absence of such systems may be principled, not accidental. If so, the gap remains unexpected and unexplained within the ABC model.[16]

5.5 Summary and conclusion

The Agreement By Correspondence (ABC) model of long-distance agreement operates on the premise that similarity is a source of correspondence between segments in output strings. Output-output correspondence is enforced by a similarity-based hierarchy of CORR-C↔C constraints. Long-distance agreement is enforced by IDENT-CC constraints that demand feature

---

[16] Rose & Walker (2004) suggest that the relative similarity of segments might be partly determined by an independent similarity metric, like that of Frisch, Pierrehumbert, & Broe (2004), and that some features may be weighted so that they contribute more to the evaluation of similarity than others. These issues are taken up in §6.1.

matching between correspondent segments. These constraints compete with faithfulness constraints of the IDENT-IO family. The interaction of CORR-C↔C, IDENT-CC and IDENT-IO constraints is able to provide a straightforward account of the similarity effects attested in retroflex consonant harmony systems among South Asian languages.

While the ABC framework is generally successful in modelling similarity effects, some outstanding issues remain. Most notably, the model predicts the possibility of consonant harmony systems in which retroflex agreement is sensitive to laryngeal features. The survey of retroflex consonant harmony in South Asia found no such system. This may be an accidental gap, but cross-linguistic typological evidence suggests otherwise. Minor place harmonies of all types, whether coronal or dorsal, are often sensitive to similarity in terms of manner/stricture features, but not in terms of laryngeal features (Rose & Walker, 2004, p. 485).

This serves to highlight an important point concerning the ABC framework: The framework is intended to model long-distance agreement and similarity effects through constraint interaction; it is not intended to serve as a means of evaluating similarity itself. The model does make some predictions concerning similarity on the basis of subset-superset relations between CORR-C↔C constraints (including the prediction that agreement for laryngeal features might be a condition for retroflex agreement). Nevertheless, it is generally acknowledged that similarity effects might also be informed by other factors such as language-specific contrast and/or the intrinsic perceptual properties of certain features. These and other issues surrounding the evaluation of similarity are the focus of the following chapter.

Chapter 6
Evaluating similarity

Retroflex consonant harmony in South Asia exhibits robust similarity effects. Interacting segments are those that share a high degree of similarity, particularly with respect to manner of articulation. This is not surprising. Similarity effects are a typological property of most long-distance segmental interactions, whether assimilatory (Hansson, 2001; 2010; Rose & Walker, 2004; Rose, 2011), or dissimilatory (Pierrehumbert, 1993; Suzuki, 1998; Frisch, Pierrehumbert, & Broe, 2004). In consonant harmony systems, pairs of segments that are highly similar interact through assimilation of some feature. As a result, they become even more similar or, in some cases, completely identical. In long-distance dissimilation, pairs of similar segments are avoided in favour of pairs of dissimilar segments. Such systems often show gradient effects: the more similar a pair of segments, the more stringently they are avoided; the less similar a pair of segments, the more freely they co-occur (Frisch, Pierrehumbert, & Broe, 2004; Coetzee & Pater, 2008).

The role of similarity in conditioning long-distance interactions is well established. However, an unresolved issue in phonological theory concerns the evaluation of similarity. Exactly how is similarity evaluated? How similar must segments be in order to interact and how is the relative similarity of any pair of segments determined? Is similarity evaluated over abstract representations consisting of phonological features or over psychoacoustic perceptual properties? Why do some features or psychoacoustic properties count more than others in determining similarity? For instance, why does manner of articulation appear to have a greater bearing on similarity than laryngeal features in retroflex consonant harmony systems? What role does contrast play in determining similarity? Is similarity evaluated only over contrastive features/properties, or is it also determined to some extent by redundant features/properties?

These and many other questions remain unresolved. The goal of the present chapter is not to provide a definitive solution to any of these issues, but simply to explore some of the hypotheses that have been proposed in the literature, and to evaluate them in light of evidence from retroflex consonant harmony in South Asia. Three different hypotheses are discussed: the natural classes similarity metric of Frisch, Pierrehumbert, & Broe (2004) (§6.1), the contrastive hierarchy model of Mackenzie (2005; 2009; 2011) (§6.2), and the Dispersion Theory account of Gallagher (2010; 2012) (§6.3). The first two hypotheses reflect different approaches to representational similarity; the third hypothesis reflects one possible approach to perceptual similarity. I identify the challenges that arise in attempting to extend each hypothesis to the domain of retroflex consonant harmony. Wherever possible, I point out potential solutions to these challenges and directions for future research.

6.1 The natural classes similarity metric

Some accounts of similarity operate on the premise that similarity is evaluated over phonological representations made up of phonological features. Intuitively, segments that share a large number of features are more similar than segments that share few or no features. However, the results of any metric that counts features (e.g., Pierrehumbert, 1993) will always be subject to the number and nature of the features assumed. Results can also vary depending on whether one assumes full specification of features on all segments or some form of underspecification.

In an effort to provide a more objective metric, Frisch (1996) and Frisch, Pierrehumbert, & Broe (2004) have proposed a similarity metric based on natural classes as opposed to features. In the natural classes similarity metric, the similarity of any pair of segments is calculated as the number of shared natural classes divided by the number of shared and non-shared natural classes, as shown in (1). This metric returns a value ranging between 0 and 1, in which 1 represents the highest degree of similarity (i.e., identity) and 0 represents the lowest possible degree of similarity (i.e., the segments do not share any natural classes).

(1) Natural classes similarity metric (Frisch, Pierrehumbert, & Broe, 2004)

$$\text{similarity} = \frac{\text{shared natural classes}}{\text{shared natural classes} + \text{non-shared natural classes}}$$

Although it relies on features to define natural classes, Frisch et. al. argue that the natural classes metric is superior to simple feature counting metrics because the addition of redundant features does not necessarily alter the results of the equation. While there is no limit to the number of features that might be applied to an inventory (at least in theory), there is always an upper limit to the number of distinct natural classes that can be defined over a given inventory. The addition of redundant features, regardless of their number, cannot increase the number of possible distinct natural classes in a phonological inventory.

Frisch et. al. apply the natural classes similarity metric to the problem of similarity-based co-occurrence restrictions in Arabic. Arabic verbal roots exhibit a dissimilatory co-occurrence restriction on consonants, traditionally analyzed in terms of the Obligatory Contour Principle (OCP), as applied to place of articulation (OCP-Place) (McCarthy, 1986; 1988; 1994). Arabic verbal roots typically contain two to four heterorganic consonants. Roots containing homorganic consonants are avoided. However, the restriction is not categorical. Homorganic consonants are tolerated to some degree if they are sufficiently dissimilar in terms of sonority and continuancy. This gradient effect is most evident in the coronal class, which contains more consonants than any of the other major place classes. Roots containing two coronal stops, two coronal fricatives or two coronal sonorants are almost categorically absent (O/E = 0.14, 0.04 and 0.06, respectively). However, roots containing a coronal stop and a coronal fricative are somewhat tolerated (O/E = 0.52) and roots containing a coronal stop or fricative with a coronal sonorant are well attested (O/E = 1.23 and 1.21, respectively).[1] Frisch et. al. demonstrate that the natural classes metric achieves a close fit to this gradient co-occurrence pattern. The strength of the co-occurrence restriction on a given pair of consonants is a function of the relative similarity of that pair. The higher the similarity score according to the metric, the lower the O/E ratio (reflecting a stronger co-occurrence restriction); the lower the similarity score, the higher the O/E ratio (reflecting a weaker co-occurrence restriction).

The natural classes similarity metric achieves a close fit to the Arabic data. Unfortunately, it does not predict the range of similarity effects found in other languages. According to the metric, all features and the classes they define contribute equally to the evaluation of similarity. This is problematic because there is evidence to suggest that some features or classes may contribute more to the evaluation of similarity than others, and that the relative contribution of a feature/class may be partly language dependent. For instance, Coetzee & Pater (2008) demonstrate that Muna (Austronesian) has a very similar co-occurrence pattern

---

[1] The O/E ratios cited here are based on C_1-C_2 and C_2-C_3 pairs in $C_1 \ldots C_2 \ldots C_3$ roots. C_1-C_3 pairs show a similar trend, but the restrictions are weaker as a result of the greater distance between the consonants.

to that of Arabic, but Muna differs from Arabic in the relative strength of certain features. Most notably, voicing has a stronger effect on co-occurrence rates in Muna than it does in Arabic. Variation of this type is not predicted by the natural classes metric without stipulating feature weights (as acknowledged by Frisch et. al., 2004, p. 204).

This problem becomes evident when the natural classes metric is extended to South Asian languages such as Kalasha. Table 35 shows the results of applying the metric to the coronal consonant inventory of Kalasha. The similarity scores in Table 35 are calculated assuming the same features that Frisch et. al. (2004: 201) assume for Arabic, with only such minor modifications as are necessary to account for differences in the Kalasha inventory (see Appendix C for details).

Table 35 Similarity of Kalasha coronals according to the natural classes metric of Frisch et. al. (2004)

	t	tʰ	d	dʰ	ʈ	ʈʰ	ɖ	ɖʰ	ʦ	ʦʰ	ʣ	ʧ	ʧʰ	ʤ	ʤʰ	ʈʂ	ʈʂʰ	ɖʐ	s	z	ʃ	ʒ	ʂ	ʐ	n	l	ɫ	r
t	1.00																											
tʰ	0.53	1.00																										
d	0.38	0.26	1.00																									
dʰ	0.26	0.38	0.61	1.00																								
ʈ	0.24	0.15	0.11	0.08	1.00																							
ʈʰ	0.15	0.24	0.08	0.11	0.49	1.00																						
ɖ	0.12	0.08	0.24	0.16	0.36	0.22	1.00																					
ɖʰ	0.08	0.12	0.16	0.25	0.22	0.36	0.54	1.00																				
ʦ	0.49	0.33	0.24	0.19	0.13	0.09	0.07	0.06	1.00																			
ʦʰ	0.34	0.50	0.19	0.25	0.09	0.13	0.06	0.07	0.57	1.00																		
ʣ	0.24	0.19	0.56	0.42	0.07	0.05	0.14	0.11	0.37	0.26	1.00																	
ʧ	0.25	0.18	0.13	0.10	0.23	0.15	0.11	0.08	0.46	0.29	0.20	1.00																
ʧʰ	0.18	0.25	0.10	0.13	0.15	0.23	0.08	0.12	0.29	0.47	0.14	0.53	1.00															
ʤ	0.14	0.11	0.29	0.22	0.11	0.08	0.24	0.17	0.21	0.15	0.48	0.35	0.23	1.00														
ʤʰ	0.11	0.14	0.23	0.30	0.09	0.12	0.18	0.26	0.15	0.22	0.33	0.24	0.37	0.59	1.00													
ʈʂ	0.13	0.09	0.07	0.05	0.45	0.28	0.20	0.15	0.21	0.14	0.10	0.44	0.27	0.19	0.13	1.00												
ʈʂʰ	0.09	0.13	0.05	0.07	0.29	0.46	0.15	0.21	0.14	0.21	0.07	0.27	0.45	0.13	0.20	0.54	1.00											
ɖʐ	0.07	0.06	0.14	0.11	0.21	0.15	0.49	0.33	0.11	0.08	0.22	0.19	0.13	0.44	0.29	0.36	0.23	1.00										
s	0.20	0.20	0.12	0.12	0.06	0.06	0.04	0.04	0.30	0.31	0.15	0.16	0.16	0.09	0.09	0.08	0.08	0.05	1.00									
z	0.12	0.12	0.23	0.24	0.03	0.03	0.06	0.07	0.15	0.15	0.32	0.08	0.08	0.17	0.17	0.04	0.04	0.09	0.40	1.00								
ʃ	0.11	0.11	0.06	0.07	0.10	0.10	0.05	0.06	0.17	0.17	0.09	0.29	0.29	0.13	0.14	0.16	0.16	0.08	0.44	0.20	1.00							
ʒ	0.07	0.07	0.13	0.13	0.05	0.05	0.11	0.11	0.09	0.09	0.18	0.13	0.13	0.29	0.31	0.08	0.08	0.16	0.22	0.45	0.37	1.00						
ʂ	0.06	0.06	0.03	0.03	0.19	0.19	0.10	0.10	0.08	0.08	0.04	0.15	0.15	0.07	0.08	0.30	0.30	0.15	0.20	0.10	0.41	0.18	1.00					
ʐ	0.04	0.04	0.06	0.07	0.10	0.10	0.21	0.21	0.05	0.05	0.09	0.07	0.07	0.15	0.16	0.14	0.14	0.31	0.11	0.20	0.18	0.40	0.37	1.00				
n	0.16	0.16	0.33	0.33	0.05	0.05	0.09	0.09	0.13	0.13	0.27	0.07	0.07	0.14	0.14	0.04	0.04	0.07	0.10	0.19	0.05	0.09	0.03	0.05	1.00			
l	0.08	0.08	0.15	0.15	0.02	0.02	0.04	0.04	0.07	0.07	0.13	0.03	0.03	0.06	0.06	0.02	0.02	0.03	0.21	0.41	0.09	0.19	0.05	0.09	0.36	1.00		
ɫ	0.04	0.04	0.08	0.08	0.05	0.05	0.09	0.09	0.03	0.04	0.06	0.02	0.02	0.03	0.03	0.04	0.04	0.07	0.10	0.19	0.05	0.09	0.11	0.23	0.17	0.41	1.00	
r	0.04	0.04	0.08	0.08	0.05	0.05	0.09	0.09	0.04	0.04	0.07	0.02	0.02	0.03	0.03	0.04	0.04	0.08	0.11	0.19	0.05	0.10	0.12	0.24	0.18	0.37	0.83	1.00

The results in Table 35 confirm that the natural classes similarity metric makes undesirable predictions about similarity effects in Kalasha. Most importantly, it predicts a greater degree of similarity between dental /t/ and palatal /ʧ/ (0.25) than between /t/ and any retroflex plosive, including /ʈ/ (0.24), /ʈ^h/ (0.15), /ɖ/ (0.12), and /ɖ^h/ (0.08). This, in turn, predicts that /t/ is more likely to harmonize with palatal /ʧ/ than with any retroflex plosive. As it is, dental plosives do *not* harmonize with palatal affricates in Kalasha, although they *do* harmonize with retroflex plosives, regardless of laryngeal distinctions. The evidence from Kalasha suggests that manner of articulation along the strident/non-strident dimension (i.e., the distinction between plosives and affricates) plays a greater role in determining similarity than other features. Dental plosives are considered more similar to retroflex plosives than to any strident segment, presumably because dental and retroflex plosives are both non-strident. Again, this prediction does not follow from the natural classes similarity metric without, perhaps, stipulating feature weights.[2]

It is worth noting that the same problem is inherent even in Frisch, Pierrehumbert, & Broe's (2004) analysis of Arabic. The authors do not apply their metric globally to the Arabic phoneme inventory. Rather, they apply it only within each major articulator class. They use it only to determine the relative similarity of consonants that share the same major place of

---

[2] Other studies have pointed out that the natural classes similarity metric can make erroneous predictions when applied to asymmetrical phoneme inventories (Hansson, 2001, pp. 435-436; 2010, pp. 330-331; Mackenzie, 2009, pp. 63-64). The Kalasha coronal inventory does contain an important asymmetry: dental and retroflex obstruents include stops, affricates and fricatives (e.g., /t, ʦ, s/ and /ʈ, ʈʂ, ʂ/), whereas palatal obstruents include only affricates and fricatives (e.g., /ʧ, ʃ/). However, the undesirable predicitons of the metric do not appear to stem from this asymmetry. The same predictions are made even when the metric is applied to a hypothetical language, Kalasha′, in which a fully symmetrical system is assumed, i.e., one that includes unaffricated palatal stops (e.g., /c/). Under these conditions the pair /t, ʧ/ (0.31) is still deemed more similar than the pair /t, ʈ/ (0.25). See Appendix C.

articulation but differ with respect to manner. Consonants that belong to different major place classes are assigned an automatic similarity score of 0. This reflects the fact that major place has a stronger effect on similarity than any other feature, a result that must be stipulated because it does not follow from the metric itself.

In sum, the natural classes similarity metric achieves a close fit to attested similarity effects in Arabic (at least within major articulator classes), but does not necessarily predict the range of similarity effects found in other co-occurrence patterns cross-linguistically. This problem might be avoided by incorporating weighted features into the metric. However, barring some non-arbitrary means of assigning feature weights, this move would only weaken the predictive power of the metric. With the introduction of weighted features, similarity would be determined in part by factors independent of the metric. Consequently, the contribution of the metric itself would be significantly reduced. The following section explores an alternative approach to representational similarity.

6.2 Similarity and the contrastive hierarchy

Cross-linguistic surveys have observed that consonant harmony and other long-distance co-occurrence restrictions are constrained by similarity; the set of interacting segments are those that are highly similar to one another in some respect (Rose & Walker, 2004; Hansson, 2010). However, there is also evidence that phonological contrast plays a role in determining the set of interacting segments. In his cross-linguistic survey of consonant harmony systems, Hansson makes the following observation:

> *"In general, the set of consonants that interact in any given consonant harmony system typically consists of those that are contrastively specified for the feature*

in question; segments that are redundantly [+F] (or redundantly [–F]) are completely inert and transparent to the harmony." (Hansson, 2010, p. 328)

Observations of this type raise questions about the relation between similarity and contrast. To what extent are similarity effects determined by contrast? Rose & Walker express doubts about any strong deterministic relation between the two, primarily on the basis of evidence from the Nilotic language, Anywa (2004, pp. 517–518). More recently, however, Mackenzie (2005; 2009; 2011) has argued that similarity *can* be determined by contrastive feature specifications, even in the case of Anywa, assuming an appropriate model of contrast.

This section reviews Mackenzie's contrast-based account of similarity in light of evidence from retroflex consonant harmony in South Asia. Section §6.2.1 introduces the contrastive hierarchy model assumed by Mackenzie. Sections §6.2.2 and §6.2.3 summarize her analysis of similarity effects in consonant harmony systems. Evidence from Indus Kohistani, a Dardic language of South Asia, is reviewed in section §6.2.4, where I demonstrate that similarity effects in that language cannot be explained in terms of Mackenzie's analysis. Finally, in section §6.2.5, I briefly sketch a modification that makes it possible to extend Mackenzie's analysis to Indus Kohistani. However, without further constraints, the model predicts (as yet) unattested similarity effects. Thus, the viability of the proposed modification remains uncertain, barring further research.

6.2.1 The contrastive hierarchy

Mackenzie (2005; 2009; 2011) adopts the contrastive hierarchy theory of contrast (Dresher, 2003; 2009). In this model, features are assigned to segments in accordance with the Successive Division Algorithm (SDA) in (2). The SDA is formulated from the perspective of the language learner, who begins with a single undifferentiated phoneme and proceeds to make a series of divisions until each phoneme has a unique representation.

(2) The Successive Division Algorithm (Dresher, 2009, p. 16)

a. Begin with no feature specifications: assume all sounds are allophones of a single undifferentiated phoneme.

b. If the set is found to consist of more than one contrasting member, select a feature and divide the set into as many subsets as the feature allows for.

c. Repeat step (b) in each subset: keep dividing up the inventory into sets, applying successive features in turn, until every set has only one member.

Contrastive feature specifications are governed by a feature hierarchy, which determines the order in which features are accessed by the SDA. Features higher in the hierarchy are accessed first and have scope over features lower in the hierarchy. Feature hierarchies are language-specific. Cross-linguistic variation in phonological behaviour is attributed to different contrastive feature specifications arising from different feature hierarchies, as applied to different inventories.

Assuming the contrastive hierarchy model, Mackenzie argues that similarity effects in consonant harmony systems can be reduced to one of the following:

(3) Interacting segments in consonant harmony systems (Mackenzie, 2009)

a. The natural class of segments contrastively specified for the harmonic feature.

b. Segments that differ only in a single marked and contrastive feature specification.

The following sections present illustrative examples of both (3)(a) and (3)(b).

6.2.2 Similarity as an effect of natural classes

Mackenzie argues that, in most cases, the set of interacting 'similar' segments in a consonant harmony system is simply the natural class of segments contrastively specified for the harmonizing feature. Within the contrastive hierarchy model, feature hierarchies are language-specific. Thus, two languages with identical or near-identical inventories might have different contrastive specifications if they make use of different hierarchies. If contrastive specifications determine similarity effects, then the model predicts that languages with identical or near-identical inventories might exhibit different similarity effects.

Mackenzie illustrates this point with reference to the Nilotic languages, Anywa and Dholuo. As shown in (4) and (5), these languages have near-identical coronal inventories. Both languages contrast dental and alveolar stops, voiced and voiceless. The only difference is the presence of prenasalized stops in Dholuo. Significantly, both languages have the same asymmetry with respect to nasal stops; in each case alveolar /n/ lacks a phonemic dental counterpart (i.e., */n̪/).

(4) Anywa coronal inventory (Mackenzie, 2009; 2011)

DENTAL	ALVEOLAR
t̪	t
d̪	d
	n

(5) Dholuo coronal inventory (Mackenzie, 2009; 2011)

DENTAL	ALVEOLAR
t̪	t
d̪	d
ⁿd̪	ⁿd
	n

Anywa and Dholuo both exhibit dental consonant harmony. Dental and alveolar stops do not co-occur within the same word. Despite their genetic relation and near-identical inventories, the two languages exhibit different patterns with respect to alveolar /n/. In Anywa, alveolar /n/ participates in harmony, surfacing as allophonic dental [n̪] whenever there is a (non-adjacent) dental stop in the word. In Dholuo, alveolar /n/ does not participate in harmony. It co-occurs freely with dental and alveolar stops and no dental allophone is reported. Representative examples from these languages are listed in (6) and (7).

(6) Dental harmony in Anywa: nasals participate (Mackenzie, 2009; 2011)

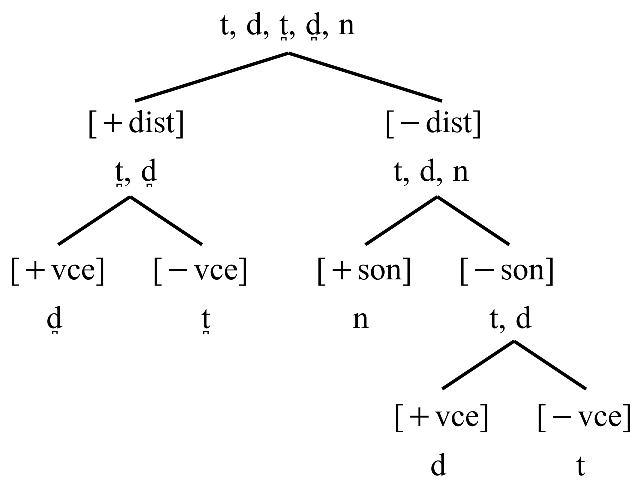
t̪ùd̪	'ropes'	tūud	'pus'
n̪ùd̪ò	'to lick'	núudó	'to press something down'
ōd̪óòn̪	'mud'	dīn	'to thrash something'

(7) Dental harmony in Dholuo: nasals do not participate (Mackenzie, 2009; 2011)

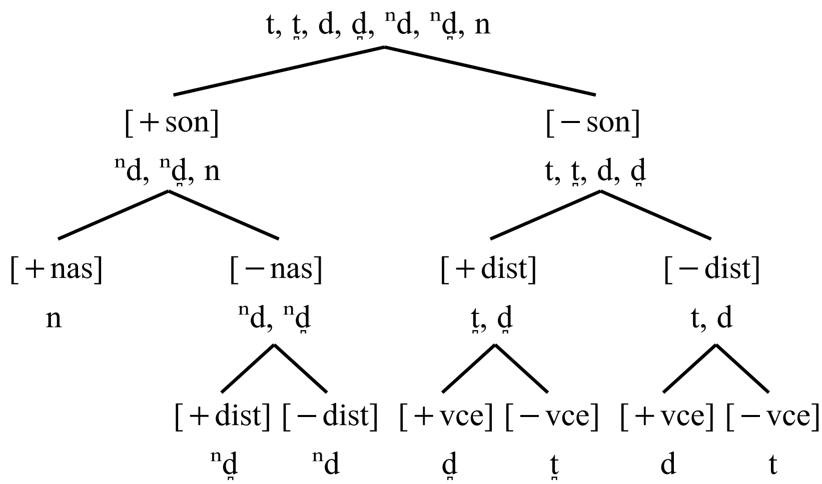
t̪ed̪o	'to forge'	tedo	'to cook'
d̪ɔd̪ɔ	'to suckle'	diedo	'to balance'
t̪uno	'breast'	dino	'deaf, to be stopped up'
t̪ʊɔn	'brave man'	tîn	'small'

The pattern in Dholuo is unsurprising. Nasals are not expected to participate in dental harmony because there is no contrast between dental and alveolar nasals. The pattern in Anywa, however, is unexpected. Nasals participate in dental harmony despite the fact that there is no contrast between dental and alveolar nasals. Rose & Walker take this as counter-evidence to the claim that contrast determines participation in consonant harmony (2004, p. 517), but Mackenzie attributes the variable behaviour of /n/ to different contrastive feature specifications arising from different contrastive hierarchies. She proposes the hierarchies in (8) and (9) for Anywa and Dholuo, respectively.

(8) Anywa contrastive hierarchy: [distributed] > [sonorant] > [voice]



(9) Dholuo contrastive hierarchy: [sonorant] > [nasal] > [distributed] > [voice]



The critical difference between (8) and (9) is the position of [distributed] relative to other features in the hierarchy. In (8) [distributed] has scope over all other features. Thus, in Anywa, all coronals are specified for some value of [distributed], including alveolar /n/, which is contrastively specified as [−dist] despite the fact that it has no [+dist] counterpart. In (9) [sonorant] and [nasal] have scope over [distributed]. Thus, in Dholuo, alveolar /n/ is not specified for any value of [distributed] because the former features suffice to represent it uniquely. The natural class of segments contrastively specified for (some value of) [distributed] includes /t, d, t̪, d̪, n/ in the case of Anywa and /t, d, ^{n}d, t̪, d̪, ^{n}d̪/ (but not /n/) in the case of Dholuo. These are precisely the classes that participate in dental harmony in each case. Thus, assuming the specifications in (8) and (9), the class of segments participating in consonant harmony can be defined in each case as the natural class of segments contrastively specified for the harmonizing feature.

While many examples of consonant harmony can be analyzed along the same lines as Anywa and Dholuo, Mackenzie (2009) acknowledges that some cannot. Some cases of consonant harmony require reference to a notion of similarity that is distinct from the notion of natural classes. In these cases, Mackenzie argues that contrast, as determined by the contrastive hierarchy, still plays a vital role. Her analysis of these cases is illustrated in the following section.

6.2.3 Similarity as an effect of minimal contrast

Wherever the class of interacting segments cannot be reduced to those that are contrastively specified for the harmonizing feature, Mackenzie argues that interacting segments are those that differ only (i.e., “minimally”) in a single marked and contrastive feature specification. This approach can be illustrated with reference to Hausa (Afro-Asiatic, Chadic). As shown in (10), Hausa obstruents exhibit laryngeal contrasts for voicing and glottalization. Glottalized consonants are realized as implosive in the case of labial and coronal stops and ejective otherwise.

(10) Hausa consonant inventory (Mackenzie, 2009)[3]

LABIAL	CORONAL	PALATAL	VELAR	LABIALIZED VELAR	PALATALIZED VELAR	LARYNGEAL
(f, fʲ)	t	ʧ	k	kʷ	kʲ	
b	d	ʤ	g	gʷ	gʲ	
ɓ	ɗ	(j')	k'	kʷ'	kʲ'	ʔ
f, fʲ	s	ʃ				h
	z					
	s'					
m	n					
	l					
	r					
	ɽ					
		j		w		

The co-occurrence of glottalized segments, implosive or ejective, is highly constrained in Hausa. Multiple glottalized segments do not co-occur within roots unless they are identical, as shown in (11)(a). In roots containing multiple heterorganic consonants, no more than one glottalized consonant can occur (11)(b), and glottalized segments cannot co-occur with their homorganic non-glottalized counterparts (11)(c).

(11) Laryngeal co-occurrence restrictions in Hausa (Mackenzie, 2009)

a. ɓaɓe ‘quarrel’

s’as’a ‘rust’

k’uk’uta ‘try hard’

---

[3] There is no contrast between /f/ and /p/. From a phonological point of view, /f/ can be regarded as the voiceless counterpart of /b/. /j’/ is a glottalized palatal glide. It is a recent innovation occurring in very few words and it is not included in Mackenzie’s (2009) discussion of laryngeal co-occurrence restrictions in Hausa.

b. *ɓ…k’, *s’…ɓ, *k’…ɗ, etc.

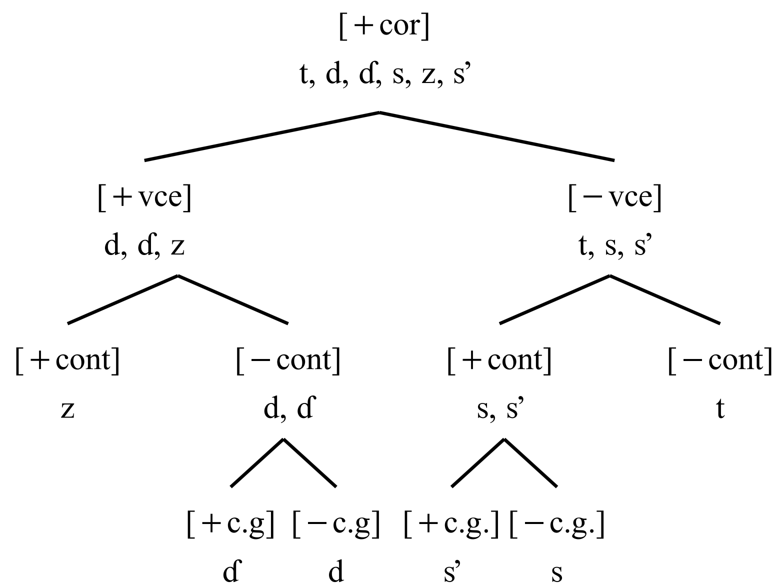
c. *ɓ…b, *s’…s, *k’…k, etc.

Co-occurrence patterns like that in Hausa have been analyzed as dissimilation with an identity exemption (MacEachern, 1997). As a general rule, there is dissimilation of the glottalization feature, as in (11)(b), but identical consonants are exempt, as in (11)(a). However, patterns of this type can also be analyzed as similarity-sensitive laryngeal harmony (Hansson, 2001; 2010). Roots containing highly similar homorganic consonants cannot disagree for the glottalization feature, as in (11)(c), and are subject to long-distance assimilation resulting in (11)(a). Roots containing dissimilar heterorganic consonants are not subject to harmony and may be subject to the Obligatory Contour Principle (OCP; Goldsmith, 1979) banning multiple instances of the glottalization feature, as in (11)(b).

If the Hausa pattern reflects laryngeal harmony, then the interacting segments cannot be defined as the natural class of segments contrastively specified for the harmonizing feature. Assuming that all glottalized segments are distinguished from their non-glottalized counterparts by means of a single feature, such as [constricted glottis], then no ranking of features can avoid the need to specify labials (/ɓ/ vs. /b/), coronals (/ɗ/ vs. /d/ and /s’/ vs. /s/) and velars (/k’/ vs. /k/, etc.) for that feature. Yet, not all of these consonants interact with each other in laryngeal harmony. The interacting segments are only those that agree in other features, such as place of articulation and voicing or manner. Thus, some notion of similarity is required, one that goes beyond the natural class of segments contrastively specified for the harmonizing feature.

In cases such as this, Mackenzie (2009) argues that interacting segments can be defined as those that differ minimally in a single marked and contrastive feature specification. In the case of Hausa, she proposes the feature hierarchy in (12). For the sake of brevity, only the portion of the hierarchy pertaining to coronal obstruents is shown here. The hierarchy in (12) yields the *marked* contrastive feature specifications in (13). In Mackenzie’s analysis, marked features are those represented by “ + ” values.

(12) Hausa contrastive hierarchy: [coronal] > [voice] > [continuant] > [constricted glottis]



(13) Marked contrastive feature specifications for Hausa coronals

t	d	ɗ	s	s’
[+cor]	[+cor]	[+cor]	[+cor]	[+cor]
	[+vce]	[+vce]	[+cont]	[+cont]
		[+c.g.]		[+c.g.]

Assuming the feature specifications in (13), interacting segments can be defined as those that differ minimally in the harmonizing feature. That is, they differ only in a single marked and contrastive feature specification, in this case [+constricted glottis]. Thus, the interacting pairs are /d, ɗ/ and /s, s’/. Other pairs, such as /t, ɗ/, /s, ɗ/, /t, s’/ and /d, s’/, do not interact because each one differs in [+voice] and/or [+continuant], in addition to [+constricted glottis].

Mackenzie's account makes an important prediction regarding consonant harmony systems with similarity effects like that in Hausa. It predicts that the output of similarity-sensitive harmony should always be complete identity of segments. If interacting segments differ only in the harmonizing feature, then agreement for that feature will always result in complete identity. This prediction is borne out in Hausa (cf. (11)(a)) and in many other cases of consonant harmony, particularly those involving laryngeal features. This observation has led others to make the same prediction on independent theoretical grounds (e.g., Gallagher & Coon, 2009). Unfortunately, as I will demonstrate in §6.2.4 below, the evidence from retroflex consonant harmony in South Asia does not support this prediction.

To summarize, Mackenzie (2005; 2009; 2011) argues that phonological similarity is evaluated over contrastive feature specifications, as determined by a contrastive hierarchy. In this approach, interacting 'similar' segments are reduced to either: (i) the natural class of segments contrastively specified for the harmonizing feature; or (ii) those segments that differ minimally in a single marked and contrastive feature specification. In the following section I review evidence from Indus Kohistani, a Dardic language of South Asia, and demonstrate that the interacting segments in consonant harmony systems cannot always be reduced to the sets identified by Mackenzie's analysis.

6.2.4 Evidence from Indus Kohistani

Some Indo-Aryan languages of the Dardic group exhibit a pattern of consonant harmony with similarity effects that cannot be attributed to either (i) the natural class of segments contrastively specified for the harmonizing feature, or (ii) those segments that differ minimally in a marked and contrastive feature specification. Moreover, contrary to the predictions of Mackenzie (2009) (and Gallagher and Coon (2009)), the output of similarity-sensitive harmony in these languages is not always complete identity.

These points can be demonstrated with reference to Indus Kohistani. The properties of consonant harmony in Indus Kohistani have already been described in §3.3.1 and §4.3.1.2. Relevant details are repeated here for ease of reference. Indus Kohistani has the coronal obstruent inventory shown in (14). Retroflexion is contrastive in plosives and in sibilant affricates and fricatives. In addition, there are laryngeal contrasts for voicing and aspiration.

(14) Coronal obstruents of Indus Kohistani (Zoller, 2005)

DENTAL		RETROFLEX		PALATAL	
t	tʰ	ʈ	ʈʰ		
d	dʰ	ɖ	ɖʰ		
ʦ	ʦʰ	ʈʂ	ʈʂʰ	ʧ	ʧʰ
s		ʂ		ʃ	
z	zʰ	ʐ	ʐʰ	ʒ	ʒʰ

Indus Kohistani exhibits a pattern of retroflex consonant harmony that is dependent on agreement for manner of articulation along the sibilant/non-sibilant dimension. Retroflex consonants do not co-occur with non-retroflex coronals of the same manner class. If there are two coronal obstruents of the same manner within a root then they must agree for retroflexion or non-retroflexion. Thus, harmony holds between two plosives, as shown in (15), and two sibilants (affricate or fricative), as shown in (16), but not between mixed pairs of plosives and sibilants, as shown in (17).

(15) Indus Kohistani: Retroflex consonant harmony between plosives

a.	toːtáː	‘butterfly’	d^{h}ʌtrìː	‘burnt food’
	tʌ̀t^{h}	‘hot; heat’	tùnd̥	‘a kind of basket’
	t^{h}atʌ́r	‘smallpox’	dùnd̥	‘a flock, herd’
	dít^{hi}	‘given’	d^{h}ʌndà^{h}	‘dealings, business’
b.	ʈʌʈúː	‘a small horse’	ɖ^{h}àːɖ̊	‘a woodpecker’
	ʈàːʈ^{h}	‘a small rug’	ʈaɳɖáṽ	‘to beat’
	ʈ^{h}aʈʌ́r	‘shallow’	ɖáːɳɖ̊	‘a stick’
	ɖíːʈ^{hi}	‘span of hand’	ɖ^{h}ã́ːɳɖ̊	‘a pond’
c.	*t...ʈ, *ʈ...t, *t...ɖ, *ɖ....t, *d...ɖ, *ɖ...d, etc. (no retroflexes with non-retroflexes)			

(16) Indus Kohistani: Retroflex consonant harmony between sibilants (affricate and fricative)

a.	ʦíʦ^{hi}	‘nipple, breast’	ʦàs	‘a pinch’
	ʦã̀z	‘soft’	sʌzúː	‘sister’s son’
	z^{h}ã́ːz	‘a branch of a holm oak’	z^{h}ʌnzéːr	‘a kind of bird’
b.	ʧiːʧʌ̀k^{h}	‘smallpox’	ʧãːʧúː	‘a dwarf’
	ʧuʃtìː	‘absorption’	ʃã̀ːʧ	‘a kind of mungo’
	ʧ^{h}ʌ̃ʒ^{i}	‘a winnowing tray’	ʃiʃáṽ	‘to dry (up)’
	ʃòːʃaː	‘decoration’	ʃʌmʃʌtáː	‘a turtle’
c.	ʈʂìʈʂ^{h}	‘grey, spotted’	ʈʂ^{h}iʈʂáṽ	‘to learn’
	ʈʂoːʂáṽ	‘to wring out’	ʈʂ^{h}ʌnʐòː	‘a curry comb’
	ʐàːʈʂ	‘a grape’	ʐʌmʈʂúː	‘a son-in-law’
	ʂìʂ	‘a head’	ʂùːʂ	‘decent, fine, proper’

d. *ʦ…ʈʂ, *ʈʂ…ʦ, *ʧ…ʈʂ, *ʈʂ…ʧ, *ʦ…ʂ, *ʂ…ʦ, *ʧ…ʂ, *ʂ…ʧ, *s…ʂ, *ʃ…ʂ, *ʂ…ʃ, etc.
(no retroflexes with non-retroflexes)

(17) Indus Kohistani: No retroflex consonant harmony between plosives and sibilants

a.	tsaʈáṽ	‘to lick’	ʈè:ʦʰ	‘a flint’
	siʈì:	‘a whistle’	ʈʰosà:	‘a fist, punch’
	taʈʂʰáṽ	‘to carve’	ʈʂʰatáṽ	‘to plaster’
	dù:ʂ	‘a sin’	sá:ɳɖ̊	‘a bull’
b.	ʧʌʈú:	‘a grater for spices’	ʧí:ɳɖ̊	‘a crack, fissure’
	ʃòʈʰ	‘a bump, swelling’	ʃʌ̀ɳɖ̊	‘barren, castrated’

In Indus Kohistani, the set of segments that interact with each other in consonant harmony is not coextensive with the natural class of segments that is contrastively specified for retroflexion. Retroflexion is contrastive in all obstruent classes, whether sibilant or non-sibilant. No ordering of features within the contrastive hierarchy can avoid the need to specify sibilant and non-sibilant coronals for the retroflex feature (e.g., [±distributed]). However, only those obstruents that agree in manner along the sibilant/non-sibilant dimension interact with each other. Obstruent pairs that disagree in manner are not subject to harmony. Thus, there is an unambiguous similarity effect that cannot be generalized to the natural class of segments contrastively specified for retroflexion.

Given the similarity-sensitive nature of consonant harmony in Indus Kohistani, we might expect the interacting segments to be those that differ minimally in the harmonizing feature, along the lines of Mackenzie’s analysis of laryngeal harmony Hausa. Under this analysis we would expect harmony only between pairs that agree in all features apart from retroflexion (e.g., /t…ʈ/, /dʰ…ɖʰ/, /ʧ…ʈʂ/, etc.) but not between pairs that disagree in laryngeal features (e.g., /t…ɖ/, /d…ɖʰ/, etc.), continuancy (e.g., /ʃ…ʈʂ/, /ʧ…ʂ/, etc.) or both (e.g., /ʒ…ʈʂ/, /ʧʰ…ʐ/, etc.). We would also expect the output of harmony to be complete identity. As it is, laryngeal features and continuancy play no role in conditioning retroflex harmony in Indus Kohistani and the output of harmony is not always complete identity. Agreement for laryngeal features is neither a condition for retroflex harmony nor a necessary output of it (e.g., /ɖíːʈʰⁱ/ 'span of hand'; /ʈaɳɖáṽ/ 'to beat'; /ʐàːʈʂ/ 'a grape'; etc.). Agreement for manner along the continuant vs. non-continuant dimension is also unnecessary, as evidenced by the fact that affricates and fricatives interact (e.g., /ʈʂoːʂáṽ/ 'to wring out'; /ʐʌmʈʂúː/ 'a son-in-law'; etc.). Interacting segments need only agree in terms of the sibilant vs. non-sibilant distinction.

Thus, contra Mackenzie (2005; 2009; 2011), evidence from Indus Kohistani indicates that the class of interacting segments in consonant harmony systems cannot always be reduced to either (i) the natural class of segments contrastively specified for the harmonizing feature, or (ii) sets of segments that differ minimally in a marked and contrastive feature specification. However, before leaving this discussion of Mackenzie's analysis, it is worth pointing out one way in which her analysis *might* be extended to patterns like that in Indus Kohistani.

6.2.5 Minimal contrast redefined

As it stands, Mackenzie's account of similarity as a product of minimal contrast cannot be extended to languages such as Indus Kohistani without predicting (erroneously) that the output of similarity-sensitive harmony should always be complete identity. However, her analysis might be extended to Indus Kohistani if we assume a slightly different approach to the notion of minimal contrast within the contrastive hierarchy, as outlined in (18).

(18) Similarity as minimal contrast redefined

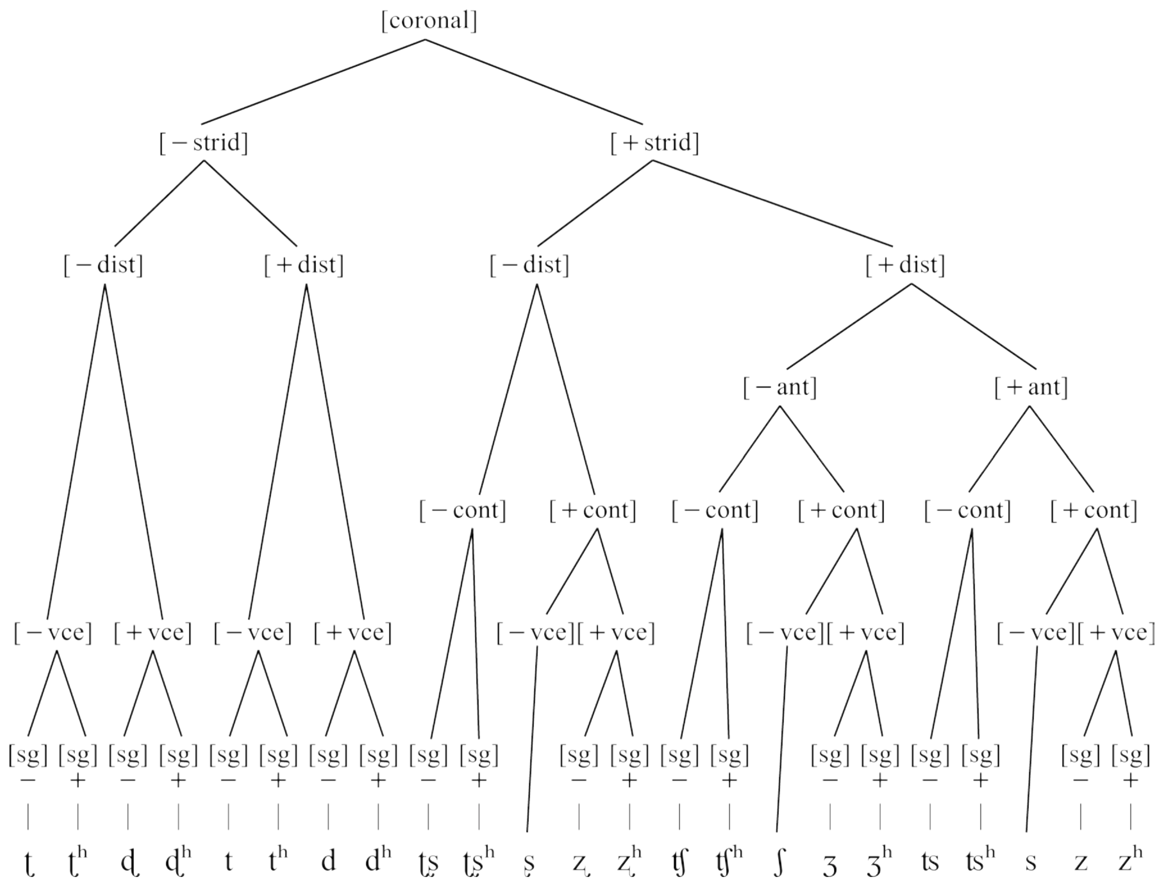
Interacting segments are those that agree with respect to all features that have scope over the harmonic feature, but may disagree with respect to:

(i) the harmonic feature itself;

(ii) any features over which the harmonic feature has scope.

In (18), 'minimal' contrast is defined only in relation to features that have scope over the harmonizing feature, not with respect to all features in the hierarchy. Interacting segments are those that agree in all features that have scope over the harmonizing feature. This leaves open the possibility that they might disagree with respect to other features, provided those features are ordered after the harmonizing feature within the contrastive hierarchy. This amendment can account for similarity effects in Indus Kohistani if we assume the contrastive hierarchy in (19).

(19) Indus Kohistani contrastive hierarchy



The crucial orderings in (19) are [strident] > [distributed], on the one hand, and [distributed] > [continuant], [voice] and [spread glottis], on the other (the place of [anterior] is discussed below). If [strident] has scope over [distributed], then harmony for the feature [−dist] (representing apicality/retroflexion) will apply only between pairs that agree with respect to stridency (i.e., plosive pairs agreeing in [−strid] or sibilant pairs agreeing in [+strid]). At the same time, if [distributed] has scope over [continuant] and all laryngeal features, including [voice] and [spread glottis], then harmony for [−dist] will not entail agreement for these other features. This would predict the attested pattern in which harmony is contingent upon agreement for stridency, but not on agreement for continuancy or laryngeal features, and the output of harmony is not necessarily identity.

Palatals may be distinguished from dentals by the feature [anterior]. If so, then [distributed] would also have scope over this feature because interacting segments do not necessarily agree with respect to anteriority; retroflex assimilation targets [+anterior] and [−anterior] segments alike (i.e., dentals and palatals). If [anterior] was ranked above [distributed], then retroflex segments would be contrastively specified as [−anterior] along with palatals. Under these conditions, we might expect retroflex assimilation to apply only to segments that agree with respect to [−anterior] (i.e., palatals and retroflexes, but not dentals). Notice, however, that retroflex segments are not contrastively specified for [anterior] in the hierarchy in (19). Rather, all retroflex segments are contrastively specified as [−distributed] and the feature [anterior] is redundant within this class. Thus, we must assume that the unfaithful mapping of [+distributed] inputs to [−distributed] outputs in retroflex consonant harmony entails a neutralization of the contrast between [+anterior] and [−anterior]. As a result, the output of retroflex assimilation does not exhibit ‘disagreement’ for [anterior], in the same way that it often exhibits disagreement for laryngeal features or [continuant].

Recall that Indus Kohistani also shows a trend toward palatal harmony. In (19), palatal harmony would be achieved through assimilation of [−anterior] and would require agreement for all features that have scope over [anterior]. The features that have scope over [anterior] include [strident] and [distributed]. Thus, under the amended definition of similarity given in (18), the hierarchy in (19) predicts that palatal harmony would only target dental affricates and fricatives, which agree with palatals in [+distributed]; it would not target retroflex segments, which are contrastively [−distributed]. The data available for Indus Kohistani (and closely related Kalasha) neither confirms nor contradicts this prediction. While Dardic languages, such as Indus Kohistani and Kalasha, show a strong tendency toward a three-way coronal agreement pattern (i.e., dental, palatal and retroflex agreement), the present study found little or no historical-comparative evidence to indicate which segments (if any) have served as the targets of palatal harmony in these languages (see discussion in §3.3.1.2). Nevertheless, it is worth noting that the predicted asymmetry is independently attested in at least one other three-way coronal harmony system. In Benchnon (a.k.a. Gimira, Afro-Asiatic), palatal /ʃ/ triggers assimilation in /s/, but not in retroflex /ʂ/. At the same time, retroflex /ʂ/ triggers assimilation in both /s/ and /ʃ/ (Hansson, 2010, p. 53, citing Rapold, 2006). This is precisely the pattern of assimilation predicted for Indus Kohistani in (19): retroflex assimilation targets both dentals and palatals, but palatal assimilation only targets dentals.

This proposal is able to account for the similarity effects in Indus Kohistani. However, without further constraints on the contrastive hierarchy, it may predict a large number of unattested similarity effects. Although the possibility of a universal feature hierarchy has been suggested (e.g., Clements, 2001), it is generally assumed that the contrastive hierarchy is subject to some degree of cross-linguistic variation. This assumption is crucial to Mackenzie's account of dental harmony in Anywa and Dholuo, where variation in the contrastive hierarchy corresponds to variation in attested patterns. If the contrastive hierarchy varies freely from language to language, we might expect to find some language in which retroflexion is ordered below voicing and/or aspiration. In such a language, retroflex harmony would be dependent on agreement for laryngeal features. As it is, the pattern in Indus Kohistani (and other South Asian languages) does not appear to be the product of chance. Cross-linguistically, laryngeal features do not appear to play a role in conditioning minor place harmonies, whether coronal or dorsal (Rose & Walker, 2004, p. 485). However, agreement for place and manner is often a condition for laryngeal harmony, with the result that the output of laryngeal harmony is often total identity (Gallagher & Coon, 2009; Mackenzie, 2009). At present, it is not clear whether these observations reflect absolute universals or mere tendencies. Either way, the patterns are unexpected unless we assume some universal or near-universal restrictions on feature ordering, such that major place and manner features (e.g., [sonorant] and [strident]) are always ordered before minor place features (e.g., [distributed] and [anterior]), which in turn are ordered before laryngeal features (e.g., [voice], [spread glottis] and [constricted glottis]). This is not to suggest that there can be no cross-linguistic variation, but only that unconstrained variation leads to the prediction of unattested similarity effects.

To summarize, Mackenzie (2005; 2009; 2011) has argued that phonological similarity is evaluated over contrastive feature specifications, as determined by a contrastive hierarchy. According to this account, interacting ‘similar’ segments in consonant harmony systems can be reduced to either (i) the natural class of segments contrastively specified for the harmonizing feature; or (ii) those segments that differ minimally in a single marked and contrastive feature specification. The Dardic languages of South Asia constitute an important counterexample. Languages such as Indus Kohistani exhibit consonant harmony with similarity effects that cannot be reduced to either of these categories. Mackenzie’s analysis might be extended to Indus Kohistani if we define minimal contrast relative to those features that have scope over the harmonizing feature within the contrastive hierarchy. However, without constraints on the possible ordering of features in the hierarchy, this model predicts harmony systems with unattested similarity effects. Research into potential universal restrictions on feature ordering is beyond the scope of the present study. Thus, this line of investigation will not be explored any further here. The following section explores a very different approach to similarity, one that appeals to perceptual factors as opposed to representational factors.

6.3 Perceptual similarity and Dispersion Theory

Despite their many obvious differences, the natural classes similarity metric (§6.1) and the contrastive hierarchy model (§6.2) share at least one thing in common: both evaluate similarity over phonological representations made up of phonological features. Recently, Gallagher (2010; 2012) has argued that long-distance laryngeal co-occurrence restrictions, including laryngeal harmony, are motivated by functional pressure to maximize the perceptual distinctness of contrasts between words. In this approach, similarity is evaluated in perceptual terms, not in terms of phonological representations.

The remainder of this chapter reviews Gallagher's proposal, beginning in §6.3.1 with the typology of laryngeal co-occurrence restrictions that it aims to explain. §6.3.2 presents Gallagher's hypothesis concerning the perceptual distinctness of laryngeal contrasts, and §6.3.3 summarizes her formal account, which is couched within the framework of Dispersion Theory (DT) (Flemming, 1995; 2004). The prospect of extending Gallagher's account to the domain of coronal co-occurrence restrictions is explored in §6.3.4. Two challenges to this prospect are identified, one concerning differences between the typologies of laryngeal and coronal co-occurrence restrictions (§6.3.4.1), and one concerning the existence of directional asymmetries within both typologies (§6.3.4.2).

6.3.1 The typology of laryngeal co-occurrence restrictions

Languages with long-distance laryngeal co-occurrence restrictions exhibit one of three patterns: dissimilation of laryngeal properties, assimilation of laryngeal properties, or a combination of both dissimilation and assimilation. These patterns are schematized in (20), where K-T represents any pair of heterorganic consonants, T-T represents any pair of homorganic consonants and the apostrophe (’) stands for any laryngeally marked segment, whether aspirate, ejective or implosive.

(20) Typology of laryngeal co-occurrence restrictions (Gallagher, 2010)

(a)	*dissimilation*	*T’-K’	T’-K	T-K’	T-K
		*T’-T’	T’-T	T-T’	T-T
(b)	*assimilation*	T’-K’	*T’-K	*T-K’	T-K
		T’-T’	*T’-T	*T-T’	T-T
(c)	*mixed*	*T’-K’	T’-K	T-K’	T-K
		T’-T’	*T’-T	*T-T’	T-T

Languages with dissimilatory restrictions avoid roots/words with multiple instances of a laryngeal property, as schematized in (20)(a). For example, Shuswap (Salishan) has a dissimilatory restriction on ejectives. Ejectives can occur with pulmonic stops, as illustrated in (21)(a), and pulmonic stops can co-occur with each other, as illustrated in (21)(b), but no root contains more than one ejective, as in (21)(c).

(21) Shuswap: Dissimilatory restriction on ejectives (Kuipers, 1974; Gallagher, 2010)

(a)	k^w’alt	‘to stagger’	T’-K, T-K’
	qet’	‘to hoist’	

(b)	k^{w}up	‘to push’	T-K
	qmut	‘hat’	
(c)	*k^w’alt’		*T’-K’
	*q’et’		

In languages with assimilatory restrictions, consonants within a root/word must agree with respect to some laryngeal feature(s), as schematized in (20)(b). For example, Chaha (Semitic) has an assimilatory restriction on ejective stops. Ejectives may co-occur within roots, as illustrated in (22)(a), and pairs of voiceless pulmonic stops may also occur, as illustrated in (22)(b), but ejectives do not occur with voiceless pulmonic stops, as in (22)(c).

(22) Chaha: Assimilatory restriction on ejectives (Rose & Walker, 2004; Gallagher, 2010)

(a)	jɨ-tʼəβkʼ	‘it is tight’	T’-K’
	jɨ-tʼəkʼɨr	‘he hides’	
(b)	jɨ-kəft	‘he opens’	T-K
	jɨ-kətf	‘it hashes (meat)’	
(c)	*jɨ-k’əft		*T’-K, *T-K’
	*jɨ-kəft’		

In languages with dissimilatory or assimilatory restrictions, heterorganic and homorganic consonant pairs behave alike. In what Gallagher terms ‘mixed’ systems, heterorganic and homorganic consonants exhibit conflicting patterns, as schematized in (20)(c). Specifically, heterorganic pairs exhibit dissimilation of laryngeal features while homorganic pairs exhibit assimilation of laryngeal features. For example, Chol (Mayan) exhibits mixed restrictions on ejectives. Pairs of heterorganic stops may include one ejective, as illustrated in

(23)(a), or no ejectives, as in (23)(c), but never two ejectives, as in (23)(d). Pairs of homorganic stops, however, always include either two ejectives, as in (23)(b), or no ejectives. Ejective stops never co-occur with their homorganic pulmonic counterparts, as in (23)(e).

(23) Chol: Mixed restriction on ejectives (Gallagher, 2010, p. 466)

(a)	p’ɨtʲ	‘to tie a load’	T’-K, T-K’
	kets’	‘obstructed’	
(b)	p’ɨp’	‘wild’	T’-T’
	ts’aʰts’	‘soak’	
(c)	ʧok	‘pull’	T-K
	patʲ	‘back’	
(d)	*p’ɨtʲ’		*T’-K’
	*k’ets’		
(e)	*p’ɨp		*T’-T, *T-T’
	*tsaʰts’		

In sum, laryngeal co-occurrence restrictions come in three varieties: dissimilatory, assimilatory and mixed. Gallagher argues that these contradictory patterns can be understood as different responses to the same functional pressure, the need to maximize the perceptual distinctness of contrasting word forms. She conducted a series of experimental studies exploring the perceptual distinctness of laryngeal contrasts. The results of these studies are summarized in the following section.

6.3.2 Perceptual distinctness of laryngeal contrasts

Gallagher points out that the typology of laryngeal co-occurrence restrictions presents a problem for most theories of markedness. The laryngeal configuration that is avoided in a language with dissimilation is precisely the configuration that is preferred in a language with assimilation, and *vice versa*. Moreover, dissimilation and assimilation can co-exist within mixed systems. Under these conditions, it is not possible to claim that a particular laryngeal configuration is universally more marked than another.

Drawing on insights from the Dispersion Theory of contrast (Flemming, 1995; 2004), Gallagher suggests that what is marked is not a particular laryngeal configuration, but rather the contrast between perceptually similar configurations. A root containing two instances of a laryngeal feature (K'-T' or T'-T') is no more or less marked than a root containing one instance of that feature (K'-T or T'-T). Rather, the contrast between these two types of roots is perceptually marked (K'-T' vs. K'-T). This approach allows her to formulate a unified account of otherwise contradictory co-occurrence restrictions. All of the laryngeal co-occurrence restrictions in Gallagher's typology neutralize the same perceptually weak contrast between one and two instances of a laryngeal specification (K'-T' vs. K'-T). Assimilatory patterns neutralize this contrast in favour of forms with two instances of the laryngeal feature (K'-T', *K'-T, *K-T'), while dissimilatory patterns neutralize the contrast in favour of forms with one instance of the feature (*K'-T', K'-T, K-T').

Gallagher proposes three hypotheses concerning the perceptual distinctness of laryngeal contrasts. These are summarized in (24), below.

(24) Hypotheses concerning perceptual distinctness of laryngeal contrasts (Gallagher, 2010)

(a) *Hypothesis 1*

Pairs of roots that contrast 2 vs. 1 instances of a laryngeal feature are less distinct than pairs of roots that contrast either 1 vs. 0 or 2 vs. 0 instances of a laryngeal feature.

(b) *Hypothesis 2*

Pairs of roots that contrast 1 vs. 0 instances of a laryngeal feature are less distinct than pairs of roots that contrast 2 vs. 0 instances of a laryngeal feature.

(c) *Hypothesis 3*

Pairs of roots with homorganic stops that contrast 1 vs. 0 instances of a laryngeal feature are less distinct than pairs of roots with heterorganic stops that contrast 1 vs. 0 instances of a laryngeal feature.

Gallagher presents experimental evidence supporting two of the three hypotheses in (24). First of all, the results of her study support the hypothesis in (24)(a), indicating that the contrast between two and one instances of a laryngeal feature (K'-T' vs. K'-T) is the most perceptually marked. This suggests that a laryngeal contrast is more difficult to perceive in the context of another laryngeally specified segment within the same root. Secondly, her experimental evidence supports the hypothesis in (24)(b), indicating that the contrast between one and zero instances of a laryngeal feature (K'-T vs. K-T) is less distinct than the contrast between two and zero instances of that feature (K'-T' vs. K-T). Although her experimental evidence does not fully support it, Gallagher also maintains the hypothesis in (24)(c), which states that the contrast between one and zero instances of a laryngeal feature is less distinct in pairs of roots containing homorganic stops (T'-T vs. T-T) than in pairs of roots containing heterorganic stops (K'-T vs. K-T). Putting all of these things together, she arrives at the hierarchy of perceptual markedness in (25), in which Δ(X:Y) means 'the perceptual distance between X and Y'.

(25) Hierarchy in the perceptual distinctness of laryngeal contrasts

2 vs. 1	<	1 vs. 0 (homorganic)	<	1 vs. 0	<	2 vs. 0
Δ([K'-T']:[K'-T])		Δ([T'-T]:[T-T])		Δ([K'-T]:[K-T])		Δ([K'-T']:[K-T])

The perceptual hierarchy in (25) is foundational to Gallagher's formal analysis of laryngeal co-occurrence restrictions. Her analysis is summarized in §6.3.3, below, and its implications for coronal co-occurrence restrictions are explored in §6.3.4.

6.3.3 Gallagher's Dispersion Theory analysis

Building on the hierarchy of perceptual distinctness in (25), Gallagher (2010) develops a formal account of laryngeal co-occurrence restrictions within the framework of the Dispersion Theory of contrast (DT) (Flemming, 1995; 2004). Unlike standard Optimality Theory (OT) (Prince & Smolensky, 2004 [1993]), in which the grammar evaluates the mapping of individual input forms to individual output forms, DT makes use of systemic markedness constraints, which evaluate the mapping between entire sets of input and output forms. This reflects the premise that the grammar evaluates contrasts between roots. The grammar constrains the range of possible roots by favouring perceptually stronger contrasts over perceptually weaker contrasts. Based on the perceptual hierarchy in (25), Gallagher proposes the systemic markedness constraints in (26).

(26) LAR(YNGEAL)DIST(ANCE) constraints (Gallagher 2010, p. 454, 458)

(a) LARDIST(2v1)-[F]

If two contrasting roots each have an [F] segment, then they do not minimally differ in [F].

(b) LARDIST(1v0)-[F]

If two contrasting roots each have two segments that may be specified for [F], then they do not minimally differ in [F].

(c) H-LARDIST(1v0)-[F]

If two contrasting roots each have two homorganic segments that may contrast for [F], then they do not minimally differ in [F].

The constraints in (26) evaluate laryngeal contrasts between roots. LARDIST(2v1)-[F] penalizes only the weakest of all laryngeal contrasts, i.e., pairs of roots that contrast two and one instances of a laryngeal feature (K'-T' vs. K'-T). LARDIST(1v0)-[F] is more general. In addition to penalizing pairs of roots that contrast two and one instances of a laryngeal feature, it also penalizes pairs that contrast one and zero instances of a laryngeal feature (K'-T vs. K-T) and pairs that differ only with respect to the position of the laryngeal feature (K'-T vs. K-T'). H-LARDIST(1v0)-[F] is a homorganic version of the same constraint; it penalizes the same set of contrasts as LARDIST(1v0)-[F] but only in roots that contain homorganic segments (T'-T' vs. T'-T, T'-T vs. T-T and T'-T vs. T-T'). There is no constraint on the contrast between two and zero instances of a laryngeal feature (K'-T' vs. K-T) because it is the strongest contrast in perceptual terms.

The systemic constraints in (26) compete with input-output faithfulness constraints of the IDENT[F] family, which penalize the neutralization of contrast that results from assimilation or dissimilation. When faithfulness to laryngeal features is undominated there are no laryngeal co-occurrence restrictions. Laryngeal co-occurrence restrictions arise only when one or more of the LARDIST constraints outrank faithfulness to laryngeal features. The rankings that account for dissimilation, assimilation and mixed patterns are outlined in (27).

(27) (a) *Dissimilation*

LARDIST(2v1)-[F] ≫ IDENT[F] ≫ LARDIST(1v0)-[F], H-LARDIST(1v0)-[F]

(b) *Assimilation*

LARDIST(1v0)-[F] ≫ IDENT[F], LARDIST(2v1)-[F], H-LARDIST(1v0)-[F]

(c) *Mixed*

LARDIST(2v1)-[F], H-LARDIST(1v0)-[F] ≫ IDENT[F] ≫ LARDIST(1v0)-[F]

Gallagher demonstrates the ‘dissimilation’ ranking of (27)(a) with reference to Shuswap. As observed earlier (see (21) in section §6.3.1), Shuswap exhibits dissimilation of ejection. Ejectives may occur initially or finally, but no root contains more than one ejective. Given an input inventory consisting of four potential contrasting root types (/k’ap’i, k’api, kap’i, kapi/) the constraint ranking in (27)(a) selects the inventory with dissimilation as the optimal candidate (/k’api, kap’i, kapi/). This is illustrated in tableau (28). Here, and in subsequent tableaus, the feature [ejective] (or [ej]) represents the contrast between ejective stops and their pulmonic counterparts.

(28) Shuswap: dissimilation in ejection

	{/k’ap’i, k’api, kap’i, kapi/}	LARDIST (2v1)-[ej]	IDENT [ej]	LARDIST (1v0)-[ej]	H-LARDIST (1v0)-[ej]
	a. {k’ap’i, k’api, kap’i, kapi}	**!		*****	
	b. {k’ap’i, k’api, kapi}	*!	*	**	
☞	c. {k’api, kap’i, kapi}		*	***	
	d. {k’ap’i, k’api}	*!	**	*	
	e. {k’ap’i, kapi}		**!		
	f. {k’api, kap’i}		**!	*	
	g. {k’api, kapi}		**!	*	
	h. {k’ap’i}		**!**		
	i. {k’api}		**!**		
	j. {kapi}		**!**		

LARDIST(2v1)-[ejective] is undominated in tableau (28). As a result, any candidate set that preserves contrast between two and one instances of [ejective] is eliminated. This includes the fully faithful candidate set in (28)(a), as well as the sets in (28)(b) and (28)(d). Notice that (28)(a) incurs two violations of this constraint, one for the pair {[k’ap’i, k’api]} and one for {[k’ap’i, kap’i]}. (28)(c) represents a system with dissimilation while (28)(e) represents one with assimilation. Both of these options avoid violations of high ranked LARDIST(2v1)-[ejective]. However, the candidate set with dissimilation in (28)(c) is preferred because it incurs fewer violations of IDENT[ejective]. The logic behind this claim is that dissimilation entails only one unfaithful mapping of [ejective], i.e., [k’ap’i] maps to either [k’api] or [kap’i]. Assimilation, however, entails two unfaithful mappings, i.e., [kap’i] and [k’api] both map to [k’ap’i]. All other candidate sets from (28)(f) to (28)(j) are eliminated for multiple violations of IDENT[ejective]. Thus, the candidate set with dissimilation in (28)(c) emerges as the winner.

Gallagher demonstrates the ‘assimilation’ ranking of (27)(b) with reference to Chaha in (29). Recall that Chaha exhibits assimilation of ejection (see examples in (22) under section §6.3.1). Ejectives always occur with other ejectives, never with voiceless pulmonic stops.

(29) Chaha: assimilation in ejection

{/k’at’i, k’ati, kat’i, kati/}	LARDIST (1v0)-[ej]	IDENT [ej]	LARDIST (2v1)-[ej]	H-LARDIST (1v0)-[ej]
a. {k’at’i, k’ati, kat’i, kati}	*!****		**	
b. {k’at’i, k’ati, kati}	*!*	*	*	
c. {k’ati, kat’i, kati}	*!**	*		
d. {k’at’i, k’ati}	*!	**	*	
☞ e. {k’at’i, kati}		**		
f. {k’ati, kat’i}	*!	**		
g. {k’ati, kati}	*!	**		
h. {k’at’i}		***!*		
i. {k’ati}		***!*		
j. {kati}		***!*		

Whereas IDENT[ejective] outranks LARDIST(1v0)-[ejective] in the dissimilation ranking of tableau (28), the inverse ranking holds for the assimilation ranking in tableau (29). Under this ranking, the candidate set with dissimilation in (28)(c) looses to the set with assimilation in (28)(e) because the former incurs violations of high ranked LARDIST(1v0)-[ejective] (i.e., one violation each for {[k’ati, kati]}, {[kat’i, kati]} and {[k’ati, kat’i]}) and the latter does not. The assimilation candidate in (28)(e) emerges as the winner, despite a double violation of IDENT[ejective], because all other candidates violate LARDIST(1v0)-[ejective] or incur more violations of IDENT[ejective].

Finally, Gallagher demonstrates the ranking responsible for ‘mixed’ systems with reference to Chol. As observed earlier (see (23) in section §6.3.1), Chol exhibits dissimilation of ejection in roots containing pairs of heterorganic stops, but assimilation of ejection when homorganic stops are involved. A single constraint ranking accounts for both of these patterns, as illustrated in tableaus (30) and (31). Crucially, IDENT[ejective] is ranked below H-LARDIST(1v0)-[ejective], which drives assimilation in homorganic pairs, but above LARDIST(1v0)-[ejective], which drives assimilation in heterorganic pairs.

(30) Chol: dissimilation in ejection (heterorganic)

{/k’at’i, k’ati, kat’i, kati/}	LARDIST (2v1)-[ej]	H-LARDIST (1v0)-[ej]	IDENT [ej]	LARDIST (1v0)-[ej]
a. {k’at’i, k’ati, kat’i, kati}	*!*			*****
☞ b. {k’ati, kat’i, kati}			*	***
c. {k’at’i, kati}			**!	

(31) Chol: assimilation in ejection (homorganic)

{/k’ak’i, k’aki, kak’i, kaki/}	LARDIST (2v1)-[ej]	H-LARDIST (1v0)-[ej]	IDENT [ej]	LARDIST (1v0)-[ej]
a. {k’ak’i, k’aki, kak’i, kaki}	*!*	****		*****
b. {k’aki, kak’i, kaki}		*!**	*	***
☞ c. {k’ak’i, kaki}			**	

In both (30) and (31), the fully faithful candidate set is eliminated by high ranked LARDIST(2v1)-[ejective]. When heterorganic stops are involved, the candidate set with dissimilation in (30)(b) is preferred, despite its many violations of low ranked LARDIST(1v0)-[ejective], because it incurs fewer violations of IDENT[ejective] than the alternative with assimilation in (30)(c). However, when homorganic stops are involved, the candidate set with assimilation in (31)(c) is preferred, despite its double violation of IDENT[ejective], because it avoids violation of high ranked H-LARDIST(1v0)-[ejective]. The dissimilation alternative in (31)(b) fails in this case because it incurs multiple violations of this constraint.

Building on the perceptual distinctness of laryngeal contrasts in (25), Gallagher is able to provide a unified account of the various and contradictory patterns in the typology of laryngeal co-occurrence restrictions. Each pattern is a response to the same functional pressure: the need to maximize the distinctness of contrasting roots by avoiding perceptually weak contrasts. The following section examines the implications of extending an analysis of this type to the domain of coronal co-occurrence restrictions involving retroflexion.

6.3.4 Extensions and implications

Gallagher's perceptual account of laryngeal co-occurrence restrictions is attractive in many respects. For one thing, it provides a unified account of seemingly contradictory patterns. For another, it is grounded in empirical observations about perceptability. The question that naturally arises is whether a similar account can be extended to long-distance co-occurrence restrictions involving other features, such as retroflexion. In the remainder of this section, I explore this possibility, noting some implications and challenges that must be addressed.

There is no doubt that the contrast between retroflex and non-retroflex coronals is perceptually marked, especially in CV contexts but also, to some degree, in VC contexts (see §1.2.3). Confusion matrices for Hindi consonants in Ahmed & Agrawal (1969) suggest that the contrast between retroflex and dental consonants in that language is more confusable than any other contrast in initial CV contexts (see also Dev, 2009). In their study, dental/retroflex confusions account for approximately 50% of all perception errors in the CV context (including errors for place, manner and laryngeal features). The same study found that place features are less confusable than laryngeal features in final VC contexts, but even in that context dental/retroflex errors still acount for the majority of place errors (approximately 62%).[4]

The results of Ahmed & Agrawal's study also show interesting parallels with the similarity effects found in retroflex consonant harmony in South Asia. For instance, their confusion matrices reveal that Hindi speakers do not generally confuse coronal plosives (Hindi /t, t^h, d, d^h, ʈ, ʈ^h, ɖ, ɖ^h/) with coronal sonorants (Hindi /n, l, r, ɳ, ɽ, ɽ^h/) in any context, although confusions within each of these categories are quite common. This is consistent with the observation in Chapter 3 that sonorants never trigger retroflex harmony in stops, and *vice versa* (with the possible exception of Nepali, where retroflex sonorants are sub-phonemic), presumably because the two classes are not sufficiently similar to interact. There are also parallels between the perceptual errors in Ahmed & Agrawal's study and the similarity effects found in the Dardic languages of South Asia. Coronal fricatives (Hindi /s, ʃ/) and affricates (Hindi /ʧ, ʧ^h, ʤ, ʤ^h/) are not generally confused with coronal plosives, although confusions within each manner class are quite common. Hindi lacks the retroflex sibilants of Dardic. Nevertheless, the trend is consistent with the observation that sibilants do not trigger harmony in plosives, and *vice versa*. Once again, this is presumably because the two classes are not sufficiently similar to interact with one another in consonant harmony.

Evidence from Ahmed & Agrawal (1969) is consistent with the hypothesis that retroflex consonant harmony may be conditioned by perceptual factors.[5] However, it is not clear that

---

[4] Ahmed & Agrawal (1969) do not discuss the conditions under which their subjects were tested. Presumably, they were tested under normal hearing conditions as opposed to noise conditions.

Gallagher's (2010) perceptual account of laryngeal co-occurrence restrictions can be extended to coronal co-occurrence restrictions in its present form. A direct adaptation of Gallagher's account would encounter at least two challenges: it would lead to questionable predictions about the typology of retroflex co-occurrence restrictions and it would not predict directional asymmetries in coronal assimilation like those found in Pengo and Kuvi (§3.1.2.2). Each of these issues is elaborated in turn below.

6.3.4.1 Typological asymmetries

The first major challenge to any extension of Gallagher's analysis lies in the fact that not all features exhibit the same typology of co-occurrence restrictions. A direct adaptation of Gallagher's account to the domain of retroflexion would entail the hypothesized hierarchy of perceptual distinctness shown in (32). The hierarchy embodies two hypotheses, which are direct adaptations of Gallagher's first two hypotheses in (24). The first states that contrast between two and one instances of retroflexion (Ṭ-Ṭ vs. T-Ṭ) is the weakest of all retroflex contrasts. The

---

[5] Perception itself may be partly influenced by the phonological contrasts in the native system of the perceiver. (Thanks to Elan Dresher for reminding me of this important point). For instance, even though native speakers of Hindi often confuse dental and retroflex stops, they are better able to perceive the dental-retroflex contrast than American English speakers. Perception studies show that adult American English speakers have great difficulty perceiving the distinction between dental and retroflex stops even after some training, a fact that has been attributed to the allophonic status of retroflexion in English (Werker, Gilbert, Humphrey, & Tees, 1981; Pruitt, Strange, Polka, & Aguilar, 1990; Pruitt, Jenkins, & Strange, 2006). Thus, if similarity is evaluated in perceptual terms, perception itself may be partly conditioned by phonological contrast (or the lack thereof). This might explain why both similarity and contrast appear to play a role in determining the class of interacting segments in consonant harmony systems. Long-distance interactions may be conditioned by similarity and similarity may be evaluated in perceptual terms, but perception may be partly conditioned by phonological contrast.

second states that contrast between one and zero instances of retroflexion (T-Ṭ vs. T-T) is perceptually weaker than contrast between two and zero instances of retroflexion (Ṭ-Ṭ vs. T-T).

(32) Hypothesized hierarchy in the perceptual distinctness of retroflex contrasts

2 vs. 1	<	1 vs. 0	<	2 vs. 0
Δ([Ṭ-Ṭ]:[T-Ṭ])		Δ([T-Ṭ]:[T-T])		Δ([Ṭ-Ṭ]:[T-T])

Recall that Gallagher proposes a third hypothesis, namely, that contrast for a laryngeal feature may be less distinct in pairs of roots containing homorganic consonants than in pairs of roots containing heterorganic consonants (Δ([T'-T]:[T-T] < Δ([K'-T]:[K-T]). There is no direct equivalent to this distinction when it comes to retroflexion. While laryngeally marked segments can be homorganic or heterorganic, retroflex segments can only be homorganic. This is because retroflexion applies only to consonants that are coronal, and therefore homorganic in the sense that they share a major articulator. However, the relation between laryngeal features and place of articulation might be analogous to the relation between coronal features and manner of articulation. That is, it might be the case that contrast for retroflexion is less distinct in pairs of roots containing consonants with the same manner of articulation than in pairs of roots containing consonants with different manners of articulation (Δ([T-Ṭ]:[T-T]) < Δ([T-Ṣ]:[T-S])).

If the hypotheses concerning retroflexion embodied in (32) are translated into systemic markedness constraints, along the lines of those in (26), and if those constraints can be freely ranked relative to faithfulness constraints penalizing the neutralization of coronal contrasts, then we predict a typology of retroflex co-occurrence restrictions parallel to the typology of laryngeal co-occurrence restrictions in (20). That is, we predict that some languages should exhibit retroflex assimilation while others exhibit retroflex dissimilation. Moreover, if we also adopt the hypothesis that contrast for retroflexion is less distinct in pairs of roots containing consonants with the same manner of articulation than in pairs of roots containing consonants with different manners of articulation, then we also predict languages with 'mixed' systems involving retroflex harmony in same manner pairs (e.g., /t…ʈ/ → [ʈ…ʈ], /s…ʂ/ → [ʂ…ʂ]) and dissimilation in different manner pairs (e.g., /ʈ…ʂ/ → [t…ʂ] or [ʈ…s]).

Assimilatory co-occurrence restrictions on retroflexion are widespread in South Asia (Chapter 3) and well-attested elsewhere (Hansson, 2001; 2010; Rose & Walker, 2004).[6] However, clear cases of retroflex dissimilation are hard to find, whether local or long-distance. Suzuki's (1998) cross-linguistic survey identifies five cases of long-distance dissimilation that *might* involve retroflexion (also cited in Alderete & Frisch, 2007). However, it is doubtful that retroflexion is the relevant dissimilating parameter in these cases. All of them involve dissimilation of rhotics, and in every case the output shows a change in manner of articulation, not necessarily a change in minor place or tongue tip orientation (e.g., /r…r/ → [n…r], [l…r] or [r…ɹ]). Dissimilation of liquids is well-attested and cases involving rhotics are probably part of the same general phenomenon.[7]

---

[6] See Table 31 (on p. 245) for a list of languages with retroflex consonant harmony based on cases reported in Hansson (2001; 2010) and Rose & Walker (2004).

[7] Hamann's (2003) cross-linguistic study of retroflexion identifies two possible cases of local retroflex dissimilation. Both of them are also doubtful. First, she reports that Proto-Dravidian */ɳʈ/ has dissimilated to [nɖ] in some Dravidian languages. Her source is Zvelebil (1970, p. 169) who lists phonological correspondences between various Dravidian languages. Following common conventions, Zvelebil's list represents [ɳɖ] as /nɖ/ in languages where the retroflexion of the nasal is limited to homorganic nasal + stop clusters, and is therefore predictable and non-phonemic. Hamann has mistakenly interpreted this as dissimilation. Secondly, Hamann reports dissimilation in the Panjabi infinitive suffix /-ɳaː/, which contains an underlying retroflex nasal that surfaces as

Those South Asian languages that do not exhibit retroflex consonant harmony exhibit a coronal co-occurrence pattern that resembles dissimilation in certain respects: they permit T-Ṭ configurations but avoid *Ṭ-Ṭ configurations. It is tempting to see in this a neutralization of the contrast between two and one instances of retroflexion (Ṭ-Ṭ vs. T-Ṭ), one that favours dissimilation. However, these languages avoid not only *Ṭ-Ṭ configurations, but also *Ṭ-T, *Ṭ-P and *Ṭ-K. Thus, the pattern in these languages is properly attributed to a historically motivated phonotactic restriction on word-initial retroflexes (see §2.3), not to dissimilation.[8]

Interestingly, Jurgec (2010, citing data from Arsenault & Kochetov, 2011) assumes a 'mixed' system for Kalasha. Recall that Kalasha (§3.3.2.1) exhibits retroflex consonant harmony in pairs of obstruents that share the same manner of articulation, whether plosives (Ṭ-

---

alveolar after stem-final retroflex consonants (e.g., /ʤaːɳ-ɳaː/ → [ʤaːɳ-naː] 'to know'). This is indeed described as dissimilation in the South Asian literature (Bhatia, 1993). However, there is another possible explanation. Recall that retroflexion is often articulated with a dynamic gesture, which Hamann (2003) calls 'flapping out', in which the place of articulation at the onset of constriction is postalveolar, but the place of articulation at the release of constriction is closer to alveolar. In view of this fact, it is possible that the /ɳ-ɳ/ sequence is not subject to dissimilation. Rather, the adjacent coronals are subject to assimilation and produced as retroflex with a 'flapping out' gesture, which results in a phonetic retroflex-alveolar sequence.

[8] It is conceivable, at least in theory, that a language with a constraint banning word-initial retroflexes could also maintain a dissimilatory constraint ranking like that in (27)(a), though the effects of the latter might be partly masked by the former. Applied to retroflexion, the constraint ranking in (27)(a) would reduce the set of possible contrasting root configurations from {Ṭ-Ṭ, T-Ṭ, Ṭ-T, T-T} to {T-Ṭ, Ṭ-T, T-T} (via elimination of *Ṭ-Ṭ). If, in addition to this, the language had an undominated constraint banning word-initial retroflexes, then this would independently eliminate *Ṭ-Ṭ and further eliminate *Ṭ-T, reducing the legal set to just {T-Ṭ, T-T}. While this scenario is conceivable, it remains highly doubtful. It is not clear that the dissimilatory constraint ranking would be learnable in a system where its effects are largely masked. The language learner would only need to establish the constraint on word-initial retroflexes in order to derive the pattern. There would be nothing to motivate the further learning of a constraint ranking whose effects cannot be independently observed.

Ṭ), affricates (Ç̣-Ç̣) or fricatives (Ṣ-Ṣ). At the same time, certain obstruent pairs that agree in retroflexion but not in manner are under-attested. These include plosive-affricate pairs (Ṭ-Ç̣/Ç̣-Ṭ, O/E = 0.13) and plosive-fricative pairs (Ṭ-Ṣ/Ṣ-Ṭ, O/E = 0.56). Jurgec has interpreted the low frequency of these pairs as an indication of retroflex dissimilation (2010, p. 366). Under this interpretation, Kalasha would qualify as a language with a mixed system, involving retroflex assimilation in same-manner pairs and retroflex dissimilation in different-manner pairs. However, the low frequency of Ṭ-Ç̣/Ç̣-Ṭ and Ṭ-Ṣ/Ṣ-Ṭ pairs does not necessarily entail unfaithful (dissimilatory) input-output mappings. Unlike the assimilation pattern, there is no historical-comparative evidence supporting dissimilation in Kalasha. As argued in §3.3.2.1, the low frequency of Ṭ-Ç̣/Ç̣-Ṭ and Ṭ-Ṣ/Ṣ-Ṭ configurations might be attributed to the *absence of retroflex assimilation* in plosive/sibilant pairs combined with the (accidental) fact that very few Ṭ-Ç̣/Ç̣-Ṭ and Ṭ-Ṣ/Ṣ-Ṭ sequences have developed from other independent sound changes.

A more plausible example of a mixed system is that of Komi-Permyak, a Finno-Ugric language of Russia. This language has coronal co-occurrence restrictions very similar to those of Kalasha (Kochetov, 2007). Obstruent pairs that agree in manner exhibit retroflex harmony, whether they are affricates (Ç̣-Ç̣) or fricatives (Ṣ-Ṣ). Obstruent pairs that disagree in manner tend to disagree in retroflexion. For instance, Ṣ-Ç̣/Ç̣-Ṣ sequences are unattested while Ṣ-Č/Č-Ṣ sequences are over-attested. Kochetov (2007) suggests that there may be historical-comparative evidence supporting dissimilation of retroflexion in different-manner pairs (e.g., ʈ͡ʂuʐ-nɨ > tɕuʐ-nɨ 'to be born'). Further research is required to determine if, in fact, dissimilation has contributed to the observed co-occurrence pattern or if other factors are at work (e.g., a tendency to neutralize contrast between retroflex and palatal affricates in favour of palatals, as noted in Kelmakov (1987)).[9] Until such a time, we may take Komi-Permyak as a possible, but unconfirmed, example of a mixed system with both retroflex assimilation and dissimilation.

Summarizing the discussion thus far, we have seen that a direct adaptation of Gallagher's account of laryngeal co-occurrence restrictions to the domain of coronal features leads to the prediction that the typology of coronal co-occurrence restrictions should include assimilatory and dissimilatory patterns, and possibly mixed patterns involving both assimilation and dissimilation. As it stands, long-distance assimilation of retroflexion is well-attested, but clear cases of dissimilation are conspicuously absent.

The absence of retroflex dissimilation may not be an accident. It would appear that different features exhibit different typologies with respect to co-occurrence restrictions. For instance, long-distance dissimilation of major place is widespread cross-linguistically, either as a categorical restriction or as a statistical trend (Suzuki, 1998; Pozdniakov & Segerer, 2007), but long-distance assimilation of major place appears to be unattested in adult grammars (Hansson, 2001; 2010; Rose & Walker, 2004; Rose, 2011). Conversely, long-distance assimilation of minor place features, coronal or dorsal, is well-attested cross-linguistically (Hansson, 2001; 2010; Rose & Walker, 2004; Rose, 2011), but long-distance dissimilation of these features is not. I know of no clear examples involving the long-distance dissimilation of minor coronal or dorsal place features. Thus, the absence of retroflex dissimilation may be part

---

[9] Thanks to Alexei Kochetov for bringing Kelmakov (1987) to my attention. It is interesting to note that, while Ṣ-Č/Č-Ṣ sequences are over-attested in Komi-Permyak, Ç-Š/Š-Ç sequences are completely unattested (Kochetov, 2007). This is the same asymmetry observed in Kalasha: retroflex fricatives occur freely with palatal affricates, but retroflex affricates do not occur with palatal fricatives (cf. discussion in §3.3.2.1, §3.3.2.2 and §4.3.1.2).

of a larger trend in which major place of articulation is often subject to dissimilation while minor place features are often subject to assimilation. Other features appear to be susceptible to either assimilation or dissimilation. In addition to laryngeal features, these include the manner features responsible for distinctions among liquids. Like laryngeal features, liquids can be subject to assimilation (Hansson, 2001; 2010; Rose & Walker, 2004; Rose, 2011) or dissimilation (Suzuki, 1998; Alderete & Frisch, 2007).[10] The source of these typological asymmetries remains an outstanding issue for phonological theory. Until they are resolved, it is not possible to provide a unified account of long-distance co-occurrence restrictions (assimilatory and dissimilatory) for all features.

6.3.4.2 Directional asymmetries

A second challenge for any extension of Gallagher's analysis is the issue of directional asymmetries. Retroflex consonant harmony is clearly regressive in South Asia. The data is ambiguous with respect to the possibility of progressive assimilation. As a result, retroflex consonant harmony may or may not be strictly regressive. However, even if it turns out to be symmetrical, with both regressive and progressive assimilation, there are independently attested directional asymmetries in other coronal harmony systems. For instance, palatal harmony in the Dravidian languages, Pengo and Kuvi, is strictly regressive despite the fact that the conditions for progressive assimilation are met (§3.1.2.2). In those languages, *T-Č configurations are avoided in favour of Č-Č configurations, via regressive palatal assimilation, but Č-T

---

[10] Even where both assimilation and dissimilation are attested, cross-linguistic trends may be evident. For instance, dissimilation appears to be more common than assimilation in laryngeal co-occurrence restrictions (Gallagher, 2012, p. 113).

configurations are preserved. They are not subject to progressive palatal assimilation (or regressive dental assimilation).

Gallagher's typology of laryngeal co-occurrence restrictions in (20) assumes that assimilatory patterns are symmetrical. They neutralize both K'-T and K-T' configurations in favour of K'-T'. The existence of directional asymmetries in consonant harmony systems entails another type of co-occurrence pattern, one that is omitted from Gallagher's typology. Symmetrical and asymmetrical assimilation patterns are schematized in (33), using C' to represent any consonant marked for a feature that might be subject to co-occurrence restrictions (e.g., laryngeal, coronal, etc.), and C to represent its unmarked counterpart.

(33) Assimilatory co-occurrence patterns[11]

(a)	*symmetrical*	C'-C'	*C'-C	*C-C'	C-C
(b)	*asymmetrical*	C'-C'	C'-C	*C-C'	C-C

The symmetrical assimilation pattern in (33)(a) is analogous to Gallagher's laryngeal assimilation pattern in (20)(a). Patterns of this type are attested as static morpheme structure constraints on coronal consonants. Alternations or historical-comparative evidence of fully symmetrical (bi-directional) assimilation are difficult to find. Sibilant harmony in Moroccan Arabic and Basque are possible examples (Hansson, 2010, pp. 157–158). The asymmetrical

[11] Another logically possible asymmetrical system would be: C'-C', *C'-C, C-C', C-C. This system would involve progressive assimilation (*C'-C → C'-C') without regressive assimilation (C-C' ↛ C'-C'). Systems of this kind appear to be very rare, although they may be attested in Bantu nasal harmony (Hansson, 2010, pp. 156–157). It is sufficient for the point at hand to establish only that *some* asymmetrical system exists. For this reason, I omit discussion of the typologically rare progressive pattern and focus on the more common regressive pattern.

pattern in (33)(b) is found in Pengo and Kuvi. Similar patterns can be found in other consonant harmony systems, including those that involve laryngeal features. For instance, Ngizim (Chadic) exhibits a directional asymmetry in voicing harmony: T-T, D-D and D-T are allowed, but *T-D is prohibited (where D represents any voiced obstruent and T represents any voiceless obstruent). Historical-comparative evidence indicates that *T-D sequences have been subject to regressive voicing assimilation: *T-D → D-D (Hansson, 2010, pp. 153–154). The asymmetrical pattern in (33)(b) may be more common than the symmetrical pattern in (33)(b), given that regressive assimilation is the norm in consonant harmony systems (Hansson, 2001; 2010).

As it stands, Gallagher's analysis cannot account for asymmetrical assimilation patterns. In Gallagher's analysis, assimilation is driven by the need to avoid a perceptually weak contrast: either the contrast between two and one instances of a feature, which includes C'-C' vs. C'-C and C'-C' vs. C-C', or the contrast between one and zero instances of a feature, which includes C'-C vs. C-C and C-C' vs. C-C. Assimilatory patterns resolve this issue by avoiding configurations containing one instance of the feature (C'-C and C-C') and preserving only those with two or zero instances of the feature (C'-C' and C-C). Asymmetrical assimilation presents a problem for this analysis because only one of the two single-feature configurations is prohibited, not both. Eliminating only *C-C', but not C'-C, does not resolve the hypothesized problem. An asymmetrical system like that in (33)(b) still preserves both of the perceptually marked contrasts: the two vs. one contrast (C'-C' vs. C'-C) and the one vs. zero contrast (C'-C vs. C-C). It is not clear how this option could ever be optimal under any ranking of systemic markedness and faithfulness constraints, like those in (26) and (27). Such a language would incur a violation of faithfulness (i.e., IDENT[F]) for eliminating *C-C' and would gain nothing in return, because the resulting system would still violate all of the systemic markedness constraints (i.e., LARDIST(2v1)-[F] and LARDIST(1v0)-[F]).

Further research is required to determine whether directional asymmetries might be derived through interaction with other independent factors, such as positional faithfulness or asymmetries in the distribution of perceptual cues to contrast. For example, it is tempting to tie the regressive direction of retroflex harmony to the asymmetry in perceptual cues to retroflexion. Contrast between retroflex and dental articulations is poorly cued in initial CV contexts, and best cued in VC contexts (§1.2.3). If assimilation for a given feature targets those positions in which contrast for the feature is least salient (Steriade, 2001), then we would expect retroflex harmony to target initial CV positions, thereby producing regressive assimilation. While this explanation seems plausible in the case of retroflexion, caution must be exercised in generalizing it to other features. The distribution of retroflex cues is unique. Many other features, including laryngeal features, are better cued in CV positions than in VC positions (Jun, 2004; Wright, 2004). By the same logic, we would expect these features to show progressive assimilation in consonant harmony. As it is, regressive assimilation represents the dominant trend in most consonant harmony systems, regardless of the features involved (with the possible exception of nasal harmony in Bantu languages, see Hansson, 2010, p. 156–157). Thus, it is not immediately clear that directional asymmetries can be derived from asymmetries in the distribution of perceptual cues in every case.

In summary, Gallagher has proposed a perceptual account of laryngeal co-occurrence restrictions within the framework of Dispersion Theory of contrast. The chief strengths of her analysis are its ability to provide a unified account of dissimilatory, assimilatory and 'mixed' co-occurrence patterns, and its grounding in empirical observations about the perceptability of laryngeal contrasts. Unfortunately, her analysis is not easily extended to other features, at least not without further refinements that are yet to be discovered. A direct extension of her analysis predicts that all features should exhibit a typology of co-occurrence restrictions parallel to that of laryngeal features, which includes assimilatory, dissimilatory and 'mixed' co-occurrence patterns. This is doubtful. Long-distance assimilation of minor place features, including retroflexion, is well-attested, but long-distance dissimilation of such features is exceedingly rare and possibly unattested. Moreover, Gallagher's analysis does not account for directional asymmetries, which are evident not only in coronal consonant harmony but also in some cases of laryngeal harmony.

6.4 Concluding remarks

Similarity effects are a typological property of most long-distance segmental interactions, whether assimilatory or dissimilatory. Retroflex consonant harmony systems in South Asia constitute an intriguing case study in similarity effects: interacting segments must agree with respect to the obstruent vs. sonorant distinction and, where applicable, to the sibilant vs. non-sibilant distinction. At the same time, they need not agree with respect to laryngeal features or the continuant vs. non-continuant distinction (e.g., affricates vs. fricatives). Phonological theory must account for patterns such as these, and explain why some features contribute to the evaluation of similarity while others contribute little or nothing.

The evaluation of similarity remains an unresolved issue in phonological theory. This chapter has reviewed three very different hypotheses concerning the evaluation of similarity. Each one faces serious challenges when extended to coronal consonant harmony systems in

South Asian languages. Further research is required before a satisfactory solution can be found. In reviewing the various proposals concerning similarity, I have endeavoured to highlight possible directions for future research. The natural classes similarity metric (§6.1) would benefit from research into the (non-arbitrary) weighting of features. The contrastive hierarchy model (§6.2) would benefit from research into possible (universal?) constraints on feature ordering. Finally, the perceptually motivated Dispersion Theory account (§6.3) would benefit from research into the source(s) of typological and directional asymmetries in long-distance co-occurrence patterns. These and other issues must be left to future research.

Chapter 7
Summary and conclusions

This dissertation has explored the nature and extent of retroflex consonant harmony in South Asia. The most important conclusions and contributions of the study are summarized in this chapter. §7.1 provides a summary of the empirical findings of the study while §7.2 provides a summary and discussion of the main theoretical contributions. Outstanding questions and directions for future research are identified where appropriate and a few closing remarks are offered in §7.3.

7.1 Empirical contributions

Long-distance co-occurrence restrictions have received little attention in the literature on South Asian languages. Retroflex consonant harmony has often gone unreported and has rarely been explored in any depth. In some cases, this has led to misleading claims about the properties of retroflex consonant harmony systems in South Asia; for instance, the claim that harmony was triggered by retroflex nasals in some Dravidian languages (e.g., Zvelebil, 1970), a claim that seems doubtful in light of the present study (see §3.1.5). Thus, two empirical goals of the present study were: (i) to document retroflex consonant harmony in as many languages as possible, and in so doing, to determine the full extent of retroflex consonant harmony in South Asia, both genetically and geographically; and (ii) to identify the typological properties of retroflex consonant harmony systems in South Asia. The following sub-sections summarize the main empirical conclusions of the study. §7.1.1 summarizes the main findings concerning the genetic and geographic distribution of retroflex consonant harmony in South Asia, while §7.1.2 summarizes the typological properties displayed in those consonant harmony systems.

7.1.1 Retroflex consonant harmony as an areal feature

The region of South Asia constitutes a linguistic area, or *Sprachbund*, in which languages of at least four distinct genetic stocks – Dravidian, Indo-Aryan, Munda and Tibeto-Burman – 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 the region is the widespread use of retroflexion as a contrastive property. A distinction between retroflex and non-retroflex coronal consonants is found in the vast majority of South Asian languages, regardless of their genetic affiliation. Although this point has been long established (e.g., Emeneau, 1956), the present study contributes supporting statistics based on a survey of descriptive literature covering almost 200 distinct South Asian language varieties (§2.1).

Cross-linguistically, retroflex consonants are commonly subject to phonotactic restrictions, and South Asian languages are no exception. The most common phonotactic restriction is the avoidance of retroflex consonants in word-initial (or other strictly pre-vocalic) positions. Historically, this restriction applied to Proto-Dravidian and Old Indo-Aryan. As a result, roots containing two non-adjacent coronal plosives were limited to just two of four logically possible configurations: T-T and T-Ṭ (with initial dentals), but not *Ṭ-T or *Ṭ-Ṭ (with initial retroflexes).[1] Most New Indo-Aryan languages, and a great many Dravidian languages, now admit retroflex plosives word-initially, although most South Asian languages continue to prohibit word-initial retroflex sonorants. Word-initial retroflex plosives have developed from a variety of sources in both Dravidian and Indo-Aryan. One of these sources is a process of

[1] In the case of Proto-Dravidian, Ṭ can be taken to represent any apical plosive, whether apico-alveolar or retroflex.

regressive long-distance retroflex assimilation. In many languages, initial dental plosives have become retroflex under the influence of a following non-adjacent retroflex plosive within the same root (T-Ṭ → Ṭ-Ṭ). As a result of this diachronic development, many South Asian languages that originally had a coronal co-occurrence pattern like that in (1)(a), with no initial-retroflexes, now exhibit a coronal co-occurrence pattern like that in (1)(b).

(1) Two co-occurrence patterns affecting dental (T) and retroflex (Ṭ) plosives

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

The co-occurrence pattern in (1)(b) can be described as retroflex consonant harmony; co-occurring coronal plosives in a root must agree in retroflexion or non-retroflexion. Using statistics calculated over lexical databases from a broad sample of South Asian languages, the present study demonstrates that this pattern occurs in many Dravidian, Indo-Aryan and Munda languages, either as a categorical restriction or as a statistical trend. Moreover, the study shows that languages with the harmony pattern in (1)(b) are concentrated in the northern half of the South Asian subcontinent, while those with the pattern in (1)(a) are concentrated in the south. Thus, a major empirical finding of the present study is that retroflex consonant harmony of the type in (1)(b) is a widespread areal trait affecting most languages in the northern half of South Asia, including languages from at least three of the four major South Asian families: Dravidian, Indo-Aryan and Munda, but not Tibeto-Burman.

Many questions remain outstanding. For instance, it is not clear how the present geographic spread of retroflex consonant harmony came to be. To what extent has retroflex consonant harmony developed independently in the various languages where it is found? To what extent has it been spread from one language to another, even across genetic lines, through multilingualism and loanword adaptation? To what extent might it be the product of substrate influences? The present study only establishes the fact that retroflex consonant harmony is found in a wide range of languages across the northern half of the South Asian subcontinent. Further research is required to determine how this state of affairs came to be, and why it is limited to some geographic regions and not others.

7.1.2 Properties of retroflex consonant harmony in South Asia

In South Asia, retroflex consonant harmony is manifested primarily (if not exclusively) as a static morpheme structure constraint (MSC), which is the result of diachronic assimilation. Historical-comparative evidence shows that assimilation was regressive, retroflex consonants were triggers, non-retroflex coronals were targets, interacting segments were constrained by similarity, and (for the most part) intervening segments were transparent. The study found a few apparent cases of long-distance retroflex assimilation in South Asia that do not fit this general description. Significantly, however, those cases that are exceptional with respect to one parameter of assimilation also tend to be exceptional with respect to other parameters. As a result, all cases of retroflex assimilation surveyed in the present study can be classified into one of two broad groups, each with a distinct set of co-occurring typological properties, as summarized in (2).

(2) Co-occurring properties in retroflex assimilation in South Asian languages

a. **Retroflex consonant harmony**	b. **Local retroflex assimilation**
Morpheme-internal	Morpheme-internal and external
Predominantly regressive	Predominantly progressive
Similarity sensitive	Not sensitive to similarity

The first and most prevalent pattern of assimilation is that which we have termed 'retroflex consonant harmony' proper. This pattern is found in Dravidian languages such as Malto (§3.1.1), Indo-Aryan languages such as Panjabi (§3.2.3), Dardic languages such as Indus Kohistani (§3.3.1) and Kalasha (§3.3.2) and Munda languages such as Mundari (§3.4), all of which exhibit the co-occurring properties in (2)(a). Exceptions to this trend include Sanskrit n-retroflexion (§3.2.1), Kalasha vowel(-consonant) harmony (§3.3.2.3), retroflex assimilation in Sherpa (§3.5), and possibly alternations in the non-past suffix of Burushaski (§3.3.5.2), all of which exhibit most (if not all) of the co-occurring properties in (2)(b). While each of these exceptions gives the appearance of long-distance retroflex assimilation, the present study has argued that they are most likely cases of local assimilation applied serially to a string of contiguous segments (cf. Gafos, 1999), not true cases of long-distance interaction.

Retroflex consonant harmony in South Asia appears to be limited to the domain of the morpheme, where it shows up as a static co-occurrence restriction on coronals. This is not a property of retroflex consonant harmony systems *per se*, but only of such systems as they occur in South Asian languages. In South Asia, retroflex consonant harmony does not produce synchronic alternations, except in the form of dialectal variation between disharmonic and harmonic forms of the same root (T-Ṭ ~ Ṭ-Ṭ). A few alternations of this kind can be found in most of the languages surveyed, but they are more numerous in data sources for some languages than for others (e.g., Gondi (Dr) in §3.1.3).

All historical-comparative evidence of long-distance retroflex assimilation points to a process of regressive assimilation in which retroflex obstruents are dominant and non-retroflex coronal obstruents are recessive (T-Ṭ → Ṭ-Ṭ). This reflects a principled and systematic avoidance of T-Ṭ configurations in favour of Ṭ-Ṭ configurations. Virtually all South Asian languages with retroflex consonant harmony have introduced word-initial retroflex plosives independent of harmony. As a result, Ṭ-P and Ṭ-K configurations, with initial retroflex stops before non-coronal stops, are typically well attested in every case. In view of this development, the widespread avoidance of Ṭ-T configurations is also significant and cannot be attributed *purely* to the historical prohibition on word-initial retroflex segments. Rather, it reflects a principled and systematic omission comparable to the avoidance of T-Ṭ configurations. Possible repair strategies for Ṭ-T configurations could include progressive retroflex assimilation (Ṭ-T → Ṭ-Ṭ) or regressive dental assimilation (Ṭ-T → T-T). The present study found no historical-comparative evidence to support either of these processes. We can only conclude that South Asian languages lacked Ṭ-T configurations historically and failed to introduce them.[2] Any conclusions about directionality or dominance that we might derive from this are purely speculative. Thus, the empirical data points clearly to regressive retroflex consonant harmony,

---

[2] Martin (2005) reports a similar phenomenon in Navajo (Athapaskan) where compound words are avoided (statistically, not categorically) if they would introduce disharmonic sibilant sequences in the lexicon. Martin suggests that Navajo speakers may opt to avoid introducing disharmonic compounds altogether, as opposed to repairing them via sibilant harmony.

but remains ambiguous with respect to the possibility of progressive retroflex assimilation, or a symmetrical system in which retroflexes can serve as both triggers and targets of assimilation.

In the retroflex consonant harmony systems of South Asia, interacting segments are highly constrained by their relative similarity in terms of manner of articulation. Synchronically, all of the languages that were found to exhibit retroflex consonant harmony exhibit a systematic co-occurrence restriction on coronal plosives within roots: they must agree for retroflexion or non-retroflexion. With only one exception, no language was found to exhibit a systematic co-occurrence restriction between plosives and sonorants, despite the fact that most of the languages surveyed distinguish one or more retroflex sonorants (e.g., /ɽ/, /ɳ/, or /ɭ/). In fact, most languages show a statistical preference for disharmonic T-Ṛ configurations over harmonic Ṭ-Ṛ configurations. This is corroborated by historical-comparative evidence. In Indo-Aryan, the lenition of single intervocalic retroflex stops typically yielded a sonorant flap ([ɽ]), which subsequently developed independent phonemic status in many languages. Wherever this has occurred, the retroflex flap has failed to trigger harmony in preceding dental plosives (e.g., Panjabi /taːɽ/ 'palm tree' < MIA /taːɖa-/). Those retroflex plosives that did trigger harmony can typically be traced to homorganic clusters or geminates, which reinforced their plosive manner and prevented lenition (e.g., Panjabi /ʈaɖɖnaː/ 'to open' < MIA /taɖɖaï/ < OIA /tardati/). A similar trend holds for Dravidian, where the triggering retroflex plosives can typically be traced to those that were members of homorganic nasal-stop clusters or geminates (§3.1.4).

A notable exception to the generalization concerning retroflex sonorants is Nepali (Indo-Aryan), where harmonic Ṭ-Ṛ configurations are preferred over disharmonic T-Ṛ configurations. Significantly, Nepali is also the only language in the survey in which the entire class of retroflex sonorants (in this case, [ɽ, ɳ]) is non-phonemic. The absence of similarity effects in Nepali may reflect the fact that phonological contrast plays a role in evaluating similarity, or otherwise determining the class of interacting segments in consonant harmony systems. A more detailed study of retroflex consonant harmony in Nepali is required. A similar pattern may also occur in some Indo-Aryan languages of the East-Central group (see footnote 34 on p. 176). This also requires further investigation. In addition, future investigations would do well to compare the system in Nepali with those of Australian languages, where retroflex harmony also appears to hold between plosives and sonorants (Hansson, 2010, p. 62).[3]

A significant contribution of the present study is the documentation of retroflex consonant harmony in Indo-Aryan languages of the Dardic group, including Indus Kohistani (§3.3.1) and Kalasha (§3.3.2). Previously reported cases of retroflex consonant harmony outside of the Dardic group include cases of harmony between coronal sibilants, on the one hand, and cases of harmony between non-sibilant coronals, on the other. However, in each of these cases, the class of interacting segments is largely coextensive with the class of segments that is contrastive for retroflexion. As a result, the role of similarity (if any) is not always evident. The Dardic languages constitute the first (and thus far, only) reported examples of retroflex consonant harmony in languages with contrastive retroflexion in both sibilant and non-sibilant coronals. They reveal striking similarity effects that are not clearly evident in other

---

[3] In some Australian languages, the pattern may be more akin to retroflex consonant-vowel harmony, with local assimilation extending over a contiguous span of consonants and vowels, as opposed to a true long-distance interaction between consonants. See examples in Dixon (2002, p. 571) and Hamann (2003, p. 123) where vowels are clearly targetted. If so, then this may explain the absence of similarity effects, as local assimilation is not necessarily conditioned by similarity.

systems: retroflex consonant harmony holds only between obstruents that agree with respect to manner along the sibilant/non-sibilant dimension. Harmony does not hold between sibilants and non-sibilants. Kalasha may show further sensitivity to the distinction between affricates and fricatives within the sibilant class, but there are asymmetries to the pattern and it may or may not be conditioned by similarity alone.

While retroflex consonant harmony in South Asia is conditioned by similarity of manner, it is not conditioned by similarity of laryngeal features. Agreement for laryngeal features, whether voicing or aspiration, is neither a condition for retroflex consonant harmony, nor a necessary consequence of it. In fact, some Indo-Aryan languages with retroflex consonant harmony may exhibit an independent dissimilatory restriction on aspiration. This appears to be the case in the Dardic languages and possibly others. However, in the present study, laryngeal co-occurrence restrictions were not explored systematically outside of word-initial $\#C_1VNC_2$ sequences containing co-occurring coronal consonants. Laryngeal co-occurrence restrictions deserve more attention in South Asia, particularly in Indo-Aryan languages, where both voicing and aspiration tend to be contrastive.

In future research, it would be interesting to explore potential interactions between coronal and laryngeal restrictions in languages with restrictions of both types. For example, Gojri has been cited as an Indo-Aryan language with a ‘mixed’ laryngeal co-occurrence restriction: laryngeal features are subject to dissimilation on co-occurring heterorganic stops, but assimilation on co-occurring homorganic stops (leading to identity). This observation was first made by MacEachern (1997) and has been cited elsewhere (e.g., Hansson, 2010; Gallagher, 2010; 2012). MacEachern’s observation was based on an analysis of data in Sharma

(1979). A cursory examination of data in the same source suggests that Gojri is among those Indo-Aryan languages with retroflex consonant harmony between coronal plosives. If so, retroflex harmony might feed laryngeal harmony in Gojri: it might create homorganic pairs which are then subject to laryngeal harmony. Alternatively, if laryngeal harmony only requires homorganicity in the broad sense (i.e., agreement for 'coronal'), then it might apply regardless of retroflex harmony. These issues warrant further investigation.

For the most part, the segments that intervene between the trigger and target of assimilation appear to be transparent in retroflex consonant harmony domains. A noteworthy case in point is Kalasha, where the vowels in retroflex harmony domains are non-retroflex despite the fact that the language distinguishes retroflex and non-retroflex vowels (§3.3.2.3). However, blocking effects cannot be ruled out completely for Kalasha and other Dardic languages. Limited data available at present suggests that sibilants may block assimilation between plosives in T-ṢṬ configurations (§3.3.4). Apart from this, no other evidence of blocking was found in any language, Indo-Aryan or otherwise. Admittedly, most of the case studies were limited to word-initial #C_1VNC_2 sequences, which might not be sufficient to reveal blocking effects, should they exist. Thus, the topic of transparency and blocking requires further investigation in South Asia. Future studies would do well to examine co-occurrence restrictions over longer domains. Phonetic studies of intervening segments in consonant harmony domains, using ultrasound or electromagnetic articulography (e.g., Walker, Byrd, & Mpiranya, 2008), might also be useful. A study of this kind would be particularly desirable in the case of Kalasha, where retroflexion is contrastive on vowels.

In sum, the present study has shown that retroflex consonant harmony is a widespread areal trait affecting most languages in the northern half of South Asia, including languages from at least three of the four major families in the region: Dravidian, Indo-Aryan and Munda. Retroflex consonant harmony in South Asia is manifested primarily (if not exclusively) as a static morpheme structure constraint (MSC), which is the product of diachronic assimilation. Historical-comparative evidence shows that assimilation was regressive, retroflex consonants were triggers, non-retroflex coronals were targets, interacting segments were constrained by similarity, and (for the most part) intervening segments were transparent. The study of these consonant harmony systems has much to contribute to phonological theories concerned with long-distance segmental interactions. The main theoretical contributions of the present study are summarized in the following section.

7.2 Theoretical contributions

The present study has focused on two theoretical issues that arise from the study of consonant harmony systems. The first issue, discussed in Chapter 4, concerns the mechanism(s) that drive assimilation in consonant harmony. The second issue, discussed in Chapter 6, concerns the role of similarity in conditioning consonant harmony, and the criteria by which languages evaluate similarity between interacting segments. The main contributions to these areas are summarized in §7.2.2 and §7.2.3, respectively. The dissertation also makes theoretical contributions related to the study of retroflex phonotactics. Although they are not central to the study of retroflex consonant harmony, these contributions are also summarized here in §7.2.1.

7.2.1 The evolution of retroflex phonotactics

As noted earlier, retroflex consonants are commonly subject to phonotactic restrictions cross-linguistically. A claim made in Chapter 2 is that phonotactic restrictions on retroflexion are a direct result of the evolution of retroflexion in a language. Nothing critical in the analysis of retroflex consonant harmony hinges upon this claim. Nevertheless, it is worth highlighting because it departs from previous studies, which argue for (possibly universal) perceptually motivated synchronic constraints on retroflexion (e.g., Hamilton, 1996; Steriade, 2001; Hamann, 2003). The present study demonstrates that the typology of retroflex phonotactic restrictions includes two contradictory patterns, which can be summarized as shown in (3).

(3) Two phonotactic restrictions on retroflexion

a. No word-initial or other strictly pre-vocalic (CV) retroflexes

b. No syllable-final or other strictly post-vocalic (VC) retroflexes

The restriction in (3)(a) is by far the most common cross-linguistically. Some form of it can be found in Dravidian, Indo-Aryan and Australian languages, among others. The restriction in (3)(b) is rare and has received little notice, although it is clearly attested in some Tibeto-Burman languages of the western Himalayas. The two patterns in (3) are contradictory: where retroflex segments are preferred in one, they are prohibited in the other, and *vice versa*.

The present study has argued that these contradictory patterns can be explained if phonotactic constraints on retroflexion are a direct result of the evolution of retroflexion in a language. Languages that prohibit word-initial and other strictly pre-vocalic (CV) retroflexes, as in (3)(a), are those that developed retroflexion through progressive assimilation from a preceding liquid or back vowel (e.g., rt/lt > ʈ; ut > uʈ). Since these historical antecedents were always post-vocalic (for independent reasons) they never produced retroflexes in word-initial or other strictly pre-vocalic positions. Languages that prohibit syllable-final or other strictly post-vocalic (VC) retroflexes, as in (3)(b), are those that developed retroflexion through regressive assimilation from a following liquid (e.g., Cr > ʈ). Since these historical antecedents were always word- or syllable-initial (again, for independent reasons) they never produced retroflexes in syllable-final or other strictly post-vocalic positions. In each case, the synchronic pattern can be seen as a natural consequence of the diachronic development of retroflexion.

This account departs from previous studies, which argue for (possibly universal) perceptually motivated synchronic constraints on retroflexion (e.g., Hamilton, 1996; Steriade, 2001; Hamann, 2003). Perceptual cues to retroflexion are strongest in VC transitions and weakest in CV transitions. Based on this observation, these studies argue that retroflexion is universally preferred in VC contexts over CV contexts. While these accounts provide a plausible explanation of the dominant pattern in (3)(a), they cannot explain the pattern in (3)(b), in which retroflex segments are restricted to precisely those environments where their cues are least salient, and prohibited in those environments where their cues are most salient.

The existence of (3)(b) within the typology of retroflex phonotactics raises doubts about the universality of perceptually motivated synchronic constraints on retroflexion, and suggests that the asymmetry in perceptual cues cannot be generalized into any kind of implicational universal, along the lines of Steriade’s (2001) law of apical contrast (discussed in §2.3.3). Nevertheless, the distribution of perceptual cues may still play an important role in the evolution of retroflex phonotactics. All things being equal, local retroflex assimilation is expected to be predominantly progressive because progressive assimilation preserves salient

VC transitions at the expense of less salient CV transitions ($VC_2C_3V > VC_2C_2V$), whereas regressive assimilation does just the opposite ($VC_2C_3V > VC_3C_3V$). Thus, liquids and back vowels are more likely to induce retroflexion in a following segment (e.g., Vrt > Vʈ) than in a preceding segment (e.g., CrV > ʈV). To the extent that they predict a natural bias toward progressive retroflex assimilation over regressive assimilation (under adjacency), perceptual asymmetries may account for the cross-linguistic frequency of the pattern in (3)(a) over the pattern in (3)(b).

In sum, when Tibeto-Burman languages are factored in, the evidence from South Asia suggests that phonotactic restrictions on retroflexion are a direct result of the evolution of retroflexion in a given language, not the result of universal synchronic constraints on retroflexion. In general terms, this conclusion is consistent with the theory of Evolutionary Phonology, which maintains that recurrent synchronic sound patterns are a direct reflection of phonetically motived diachronic sound changes, as opposed to innate or universal phonological knowledge (Blevins, 2004; 2006). However, the claims of the present study do not exclude the possibility of synchronic constraints governing some aspects of retroflex phonotactics. There is no reason to assume that language learners do not form synchronic phonotactic constraints based on the patterns they observe in the course of language acquisition. The claim made here is only that, where they do occur, synchronic constraints governing the distribution of retroflex segments in pre- and post-vocalic positions do not reflect innate or universal properties of human language. Rather, they reflect language-specific diachronic developments.

7.2.2 Two mechanisms of assimilation

Turning to the topic of consonant harmony, a central question in phonological theory concerns the mechanism or mechanisms responsible for assimilation in consonant harmony systems. Recent cross-linguistic studies of consonant harmony have argued that not all forms of assimilation are products of the same assimilatory mechanism. While most assimilation can be attributed to the mechanism of *feature spreading* or *gesture extension*, at least some cases of consonant harmony are the products of a distinct mechanism, known as *feature agreement* (Hansson, 2001; 2010; Rose & Walker, 2004). The distinction between agreement and spreading is motivated largely on typological grounds: consonant harmony systems tend to exhibit unique typological properties that set them apart from other assimilation patterns. This trend is unexpected if all assimilation is the product of a single assimilatory mechanism.

The present study found that retroflex assimilation patterns in South Asia are largely consistent with the typological distinction between feature agreement and feature spreading, and that they provide support for the hypothesis that consonant harmony is the product of agreement, not spreading. As noted above, all retroflex assimilation patterns in South Asia can be classified into one of two broad types, each with a distinct set of co-occurring typological properties (see (2) in §7.1.2). One of these types, which we have labelled 'retroflex consonant harmony' proper, exhibits properties consistent with feature agreement: interacting segments are constrained by similarity; assimilation is predominantly regressive; and intervening segments are transparent (for the most part). Cases of retroflex assimilation that are exceptional with respect to one of these typological properties also tend to be exceptional with respect to others, and tend to exhibit typological properties more in keeping with local feature spreading.

The typological distinction between long-distance retroflex assimilation via agreement and local retroflex assimilation via spreading is particularly evident in languages such as Kalasha, where both patterns co-exist. In Kalasha, long-distance retroflex assimilation is predominantly regressive and highly constrained by similarity. At the same time, local retroflex assimilation between adjacent segments, whether consonants or vowels, is predominantly progressive and largely unconstrained by similarity (§4.3).

The distinction between long-distance and local retroflex assimilation in South Asia appears to be most robust with respect to directionality and similarity effects. However, the distinction between the two is much weaker with respect to opacity effects. The present study found only two examples that can be construed as blocking. One is found in Sanskrit n-retroflexion, a pattern which otherwise exhibits properties consistent with local feature spreading (§3.2.1.1). The other is found in retroflex consonant harmony in the Dardic languages, a pattern which otherwise exhibits properties consistent with non-local agreement (§3.3.4). Thus, the presence or absence of opacity effects may not be a reliable indicator of the mechanism of assimilation that is at work. This is not entirely surprising. Other recent studies have also highlighted the possibility of blocking effects in consonant harmony systems (e.g., Hansson, 2007), leading Hansson (2010) to concede that "while exceedingly rare, segmental opacity effects are attested in consonant harmony and must be contended with" (p. xii).

In the interest of exploring transparency and opacity effects in South Asia, it might be useful to conduct phonetic studies of intervening segments in consonant harmony domains using ultrasound or electromagnetic articulography (e.g., Walker, Byrd, & Mpiranya, 2008). However, it should be noted that studies of this kind are limited in what they can tell us about the mechanisms of assimilation. They can certainly tell us how speakers implement words with harmony patterns, but they do not necessarily tell us anything about the mechanism or mechanisms that give rise to harmony in the first place. For instance, if retroflexion is manifested on intervening vowels in retroflex consonant harmony domains, this does not necessarily mean that the mechanism of assimilation is local feature spreading. It is conceivable that harmony could be achieved through long-distance consonant agreement, but once established, speakers may opt to sustain retroflexion over intervening segments for ease of articulation. Conversely, the absence of retroflexion on intervening vowels in consonant harmony domains cannot be taken as proof that harmony was achieved through long-distance agreement. It might have been achieved through local feature spreading, but once established, speakers may opt to leave retroflexion unpronounced on vowels or other intervening segments where its presence may be deemed phonologically redundant. Thus, in the end, similarity effects and directionality may remain the most reliable parameters distinguishing feature agreement and feature spreading.

7.2.3 The role and evaluation of similarity

Of all the typological properties of retroflex consonant harmony in South Asia, the similarity effects present the most intriguing challenge for phonological theory. Phonological theory must account for why certain properties or features condition long-distance interactions, while others do not. In the case of retroflex consonant harmony in South Asia, phonological theory must explain why interacting segments must be obstruents, and why they must agree for manner of articulation along the sibilant vs. non-sibilant dimension. At the same time, it must explain why other manner distinctions, such as the distinction between continuant and non-continuant sibilants in Indus Kohistani, do not necessarily condition harmony. Moreover, phonological theory must explain why laryngeal features never play a role in conditioning retroflex consonant harmony, or any other coronal harmony. The present study surveyed several theoretical models that aim to account for similarity effects. All of them encounter complications when extended to these particulars. The study does not offer a definitive solution to this problem, but it does have much to contribute with respect to both the role and the evaluation of similarity. These contributions are summarized in this section.

To begin with, the study of retroflex consonant harmony in South Asia supports the hypothesis that similarity does in fact play a role in conditioning long-distance interactions. This is significant given that other factors, such as contrast, also appear to play a role. The evidence from South Asia is consistent with the hypothesis that consonant harmony operates only over segments that are contrastively specified for the harmonizing feature (Mackenzie, 2009; 2011). However, participation in retroflex consonant harmony cannot always be reduced to effects of contrast alone. In the Dardic languages of South Asia, retroflexion is contrastive in sibilant and non-sibilant coronals, but interaction is limited to pairs of consonants that agree with respect to the sibilant vs. non-sibilant distinction. In this case, the class of interacting segments is not co-extensive with the class of segments that are contrastive for retroflexion. Thus, some notion of similarity is required, one that goes beyond the natural class of segments contrastively specified for the harmonizing feature (§6.2.4).

Secondly, interacting similar segments cannot always be reduced to those that differ minimally in single marked and contrastive feature specification (§6.2.4). This hypothesis, proposed by Mackenzie (2009), leads to the prediction that the output of similarity-sensitive harmony is always identity, a claim that is made independently by Gallagher & Coon (2009).

This claim has been made in the context of studies focusing on laryngeal co-occurrence restrictions, where consonant harmony often does produce identity. However, identity cannot be construed as a necessary output of all consonant harmony systems that exhibit similarity effects. Retroflex consonant harmony in South Asia is clearly conditioned by similarity of manner, but co-occurring segments that agree in retroflexion can disagree in laryngeal features, and in some cases, manner features such as continuancy (e.g., Indus Kohistani in §3.3.1). Thus, while retroflex consonant harmony in South Asia is clearly sensitive to similarity, it does not necessarily entail identity, contrary to the expectations of some theories.

Thirdly, not all features contribute equally to the evaluation of similarity. The most significant challenge for the phonological theories examined in the present study is the fact that laryngeal features play no role in conditioning retroflex consonant harmony, while some manner features do. For example, in the natural classes similarity metric of Frisch, Pierrehumbert, & Broe (2004), all phonological features and the classes they define contribute equally to the evaluation of similarity. For this reason, the metric makes undesirable predictions about similarity effects in Kalasha (§6.1). Some models that are less restrictive can provide a straightforward account of similarity effects in South Asian languages. However, they inevitably predict unattested similarity effects. For instance, the Agreement-by-Correspondence (ABC) model of long-distance assimilation is able to account for similarity effects in South Asian languages, as demonstrated in Chapter 5. However, without further restrictions, this model also predicts the possibility of retroflex consonant harmony systems in which harmony is conditioned by similarity of laryngeal features (§5.4.3), a pattern that remains unattested. Likewise, when Mackenzie's (2009) contrastive hierarchy approach is amended to account for similarity effects in Indus Kohistani, the same prediction follows (§6.2.5). Thus, a major outstanding issue for phonological theory is the fact that certain features condition retroflex consonant harmony, while others do not.

The role of perception in determining similarity effects deserves further attention. As observed in §6.3.4, confusion matrices for Hindi consonants in Ahmed & Agrawal (1969) show interesting parallels with the similarity effects found in retroflex consonant harmony systems in South Asia. For instance, confusion between consonants of the same manner class is common, but confusion between consonants of different manner classes, such as obstruents and sonorants or sibilants and non-sibilants, is not. This suggests that similarity effects may be conditioned by perceptual factors to some degree.[4]

A study of the perceptual distinctness of retroflex contrasts, along the lines of Gallagher's (2010; 2012) study of laryngeal contrasts, would be desirable. Such a study would explore the hypothesis that contrast between two and zero instances of retroflexion in a root (Ṭ-Ṭ vs. T-T) is more salient than contrast between one and zero instances of retroflexion (T-Ṭ vs. T-T), which in turn is more salient than contrast between two and one instances of retroflexion (Ṭ-Ṭ vs. T-Ṭ). It might also explore the possibility that contrast for retroflexion is more salient in pairs of roots containing consonants with different manners of articulation ([T-Ṣ] vs. [T-S]) than in pairs of roots containing consonants with the same manner of articulation ([T-Ṭ] vs. [T-T]). While a perceptual study along the lines of Gallagher (2010; 2012) is certainly desirable, it must be remembered that Gallagher's Dispersion Theory (DT) model makes some undesirable predictions with respect to long-distance coronal co-occurrence restrictions. In particular, it

---

[4] Perception, in turn, may be partly influenced by the system of contrasts in the native language of the perceiver. See footnote 5 on page 361.

predicts the possibility of languages with long-distance retroflex dissimilation, which appear to be unattested (§6.3.4.1). Moreover, it does not predict directional asymmetries, which are attested in some coronal harmony systems (§6.3.4.2). Thus, any future investigation modeled after Gallagher (2010; 2012) must also address these issues.

In sum, the study of retroflex consonant harmony in South Asia has much to contribute to phonological theories concerned with long-distance segmental interactions. It provides support for the typological distinction between feature agreement and feature spreading, and for the hypothesis that consonant harmony is the product of the former, not the latter. It also supports the hypothesis that long-distance interactions are conditioned by similarity, and provides evidence that similarity cannot be reduced to an effect of contrast alone, although contrast may be a contributing factor in the evaluation of similarity. Moreover, it indicates that not all features contribute equally to the evaluation of similarity, a fact that presents an intriguing and unresolved puzzle for phonological theory.

7.3 Concluding remarks

The subject of consonant harmony has received little prior attention in South Asia. References to it are few and fleeting, and many cases have gone unnoticed, or at least unreported, in the literature. The present study has aimed to fill this gap by providing the first in depth study of retroflex consonant harmony in South Asia. It is certainly not the final word on the subject. If nothing else, the study reveals that consonant harmony constitutes a rich and largely untapped domain of research in the region, one that has much to contribute to our knowledge of phonological interactions, both empirically and theoretically. It is hoped that the present study will promote a greater awareness of consonant harmony and stimulate further research into this previously neglected area.

Bibliography

Abbi, A. (1992). *Reduplication in South Asian languages: An areal, typological and historical study.* New Delhi: Allied Publishers.

Abbi, A. (2006). *Endangered languages of the Andaman Islands.* Muenchen: Lincom Europa.

Abraham, P. T. (1985). *Apatani grammar.* Mysore: Central Institute of Indian Languages.

Acharya, J. (1991). *A descriptive grammar of Nepali and an analyzed corpus.* Washington, D.C.: Georgetown University Press.

Acharya, K. P. (1975). *Lotha phonetic reader.* Mysore: Central Insitute of Indian Languages.

Acharya, K. P. (1983). *Lotha grammar.* Mysore: Central Institute of Indian Languages.

Ahmed, R., & Agrawal, S. S. (1969). Significant features in the perception of (Hindi) consonants. *Journal of the Acoustical Society of America, 45* (3), 758-763.

Alderete, J. D., & Frisch, S. A. (2007). Dissimilation in grammar and the lexicon. In P. de Lacy (Ed.), *The Cambridge handbook of phonology* (pp. 379–398). Cambridge: Cambridge University Press.

Alexander, & Hannah. (2000). *Phonemic summary of Bondo language.* Unpublished manuscript. Lamtaput, Orissa: Asha Kiran Society.

Allen, N. J. (1975). *Sketch of Thulung grammar with three texts and a glossary.* Ithaca: Cornell University.

Allen, W. S. (1951). Some prosodic aspects of retroflexion and aspiration in Sanskrit. *Bulletin of the School of Oriental and African Studies, University of London, 13* (4), 939-946.

Almeida, M. (1985). *A description of Konkani.* Doctoral dissertation. Washington, DC: Georgetown University.

Anderson, G. D. (1997). Burushaski phonology. In A. S. Kaye (Ed.), *Phonologies of Asia and Africa* (Vol. 2, pp. 1021–1041). Winona Lake, Indiana: Eisenbrauns.

Anderson, G. D. (2007). Burushaski morphology. In A. S. Kaye (Ed.), *Morphologies of Asia and Africa* (pp. 1233–1275). Winona Lake, Indiana: Eisenbrauns.

Anderson, G. D. (2008a). Gtaʔ. In G. D. Anderson (Ed.), *The Munda languages* (pp. 682–763). London & New York: Routledge.

Anderson, G. D. (2008b). Introduction to the Munda languages. In G. D. Anderson (Ed.), *The Munda languages* (pp. 1–10). London & New York: Routledge.

Anderson, G. D., & Harrison, K. D. (2008a). Remo (Bonda). In G. D. Anderson (Ed.), *The Munda languages* (pp. 557–632). London & New York: Routledge.

Anderson, G. D., & Harrison, K. D. (2008b). Sora. In G. D. Anderson (Ed.), *The Munda languages* (pp. 299–380). London & New York: Routledge.

Anderson, G. D., Osada, T., & Harrison, K. D. (2008). Ho and the other Kherwarian languages. In G. D. Anderson (Ed.), *The Munda languages* (pp. 195–255). London & New York: Routledge.

Anderson, G. D., & Rau, F. (2008). Gorum. In G. D. Anderson (Ed.), *The Munda languages* (pp. 381–433). London & New York: Routledge.

Andvik, E. (2003). Tshangla. In G. Thurgood, & R. J. LaPolla (Eds.), *The Sino-Tibetan languages* (pp. 439–455). London & New York: Routledge.

Annamalai, E., & Steever, S. B. (1998). Modern Tamil. In S. B. Steever, *The Dravidian languages* (pp. 100–128). London & New York: Routledge.

Aoki, H. (1968). Toward a typology of vowel harmony. *International Journal of American Linguistics, 34* (2), 142–145.

Apte, M. L., & Pattanayak, D. P. (1967). *An outline of Kumauni grammar.* Durham: Duke University.

Apte, V. S. (1957-1959). *Revised and enlarged edition of Prin. V. S. Apte's The practical Sanskrit-English dictionary.* Poona: Prasad Prakashan. Online URL: http://dsal.uchicago.edu/dictionaries/apte/.

Arokianathan, S. (1987). *Tangkhul Naga grammar.* Mysore: Central Institute of Indian Languages.

Arsenault, P. (2008). *Coronal features and retroflexion in Indo-Aryan languages.* Unpublished Generals Paper, University of Toronto.

Arsenault, P. (2009a). *Coronal consonant harmony in Indus Kohistani.* Paper presented at the MOT Phonology Workshop, University of Toronto, February 27–March 1, 2009.

Arsenault, P. (2009b). Coronal features and retroflexion in Dhivehi and other Indo-Aryan languages. *Toronto Working Papers in Linguistics, 30*, 17–33.

Arsenault, P., & Kochetov, A. (2009). *Retroflex (consonant) harmony in Kalasha.* Paper presented at the 83rd Annual Meeting of the Linguistic Society of America, San Francisco.

Arsenault, P., & Kochetov, A. (2011). Retroflex harmony in Kalasha: agreement or spreading? In S. Lima, K. Mullin, & B. Smith (Ed.), *Proceedings of the North East Linguistic Society 39* (pp. 55–66). Amherst: GLSA.

Asher, R. E. (2008). Language in historical context. In B. B. Kachru, Y. Kachru, & S. N. Sridhar (Eds.), *Language in South Asia* (pp. 31–48). Cambridge: Cambridge University Press.

Asher, R. E., & Kumari, T. C. (1997). *Malayalam.* London & New York: Routledge.

Aze, R., & Aze, T. (1973). Parengi texts. In R. L. Trail (Ed.), *Patterns in clause, sentence, and discourse in selected languages of India and Nepal, part 3* (pp. 213–362). Norman: Summer Institute of Linguistics and the University of Oklahoma.

Baart, J. L. (1997). *The sounds and tones of Kalam Kohistani, with wordlist and texts.* Islamabad & Horsleys Green: National Institute of Pakistan Studies, Quaid-i-Azam University and Summer Institute of Linguistics.

Baart, J. L. (2004). Contrastive tone in Kalam Kohistani. *Linguistic Discovery, 2* (2), 1–20.

Bailey, T. G. (1924). *Grammar of the Shina (Ṣiṇā) language.* London: Royal Asiatic Society.

Baković, E. (2000). *Harmony, dominance and control.* Doctoral dissertation. Rutgers University.

Balakrishnan, R. (1976). *Phonology of Kodagu with vocabulary.* Annamalainagar: Annamalai University.

Balakrishnan, R. (1977). *A grammar of Kodagu.* Annamalainagar: Annamalai University.

Baruah, P. N., & Bapui, V. L. (1996). *Hmar grammar.* Mysore: Central Institute of Indian Languages.

Bashir, E. (2003). Dardic. In G. Cardona, & D. Jain, *The Indo-Aryan languages* (pp. 818–894). New York & London: Routledge.

Berger, H. (1974). *Das Yasin-Burushaski (Werchikwar): Grammatik, Texte, Wörterbuch.* Wiesbaden: Otto Harrassowitz.

Berger, H. (1998a). *Die Burushaski-Sprache von Hunza und Nagar, Teil I: Grammatik.* Wiesbaden: Harrassowitz Verlag.

Berger, H. (1998b). *Die Burushaski-Sprache von Hunza und Nagar, Teil III: Wörterbuch.* Wiesbaden: Harrassowitz Verlag.

Bhaduri, M. B. (1983 [1931]). *A Mundari-English dictionary.* New Delhi: Northern Book Centre.

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

Bhaskararao, P. (1998). Gadaba. In S. B. Steever, *The Dravidian languages* (pp. 328–355). London & New York: Routledge.

Bhat, D. N. S. (1967). *Descriptive analysis of Tulu.* Poona: Deccan College Postgraduate and Research Institute.

Bhat, D. N. S. (1971). *The Koraga language.* Poona: Deccan College Postgraduate and Research Institute.

Bhat, D. N. S. (1973). Retroflexion: an areal feature. *Working Papers on Language Universals, 13*, 27-67.

Bhat, D. N. S. (1974). Retroflexion and retraction. *Journal of Phonetics, 2*, 233–237.

Bhat, D. N. S. (1998). Tulu. In S. B. Steever, *The Dravidian languages* (pp. 158–177). London & New York: Routledge.

Bhat, D. N. S., & Ningomba, M. S. (1997). *Manipuri grammar.* München: Lincom Europa.

Bhat, R. (1987). *A descriptive study of Kashmiri.* Delhi: Amar Prakashan.

Bhatia, T. K. (1993). *Punjabi: A cognitive-descriptive grammar.* London & New York: Routledge.

Bhatt, S. L. (1971). *A grammar of Tulu: A Dravidian language.* Doctoral dissertation. University of Wisconsin.

Bhattacharya, K. (1988). *Bengali phonetic reader.* Mysore: Central Institute of Indian Languages.

Bhattacharya, P. C. (1977). *A descriptive analysis of the Boro language.* Gauhati: Gauhati University.

Bhattacharya, S. (1957). *Ollari: A Dravidian speech.* Delhi: Department of Anthropology, Government of India.

Bhattacharya, S. (1975). *Studies in comparative Munda linguistics.* Simla: Indian Institute of Advanced Study.

Bickel, B. (2003). Belhare. In G. Thurgood, & R. J. LaPolla (Eds.), *The Sino-Tibetan languages* (pp. 546–570). London & New York: Routledge.

Bielenberg, B., & Nienu, Z. (2001). Chokri (Phek dialect): phonetics and phonology. *Linguistics of the Tibeto-Burman Area , 24* (2), 85–122.

Bieri, D., & Schulze, M. (1969). *Sunwar phonemic summary.* Tibeto-Burman Phonemic Summaries, 6. Kirtipur: Summer Institute of Linguistics and Tribhuvan University.

Bieri, D., & Schulze, M. (1970). Sunwar segmental synopsis. In A. Hale, & K. L. Pike (Eds.), *Tone systems of Tibeto-Burman languages of Nepal 1: Studies on tone and phonological segments* (pp. 328–44). Urbana: University of Illinois Press.

Biligiri, H. S. (1965). *Kharia: Phonlogy, grammar and vocabulary.* Poona: Deccan College Postgraduate and Research Institute.

Biswas, S. (2000). *Samsad Bengali-English dictionary* (3rd ed.). Calcutta: Sahitya Samsad. Online URL: http://dsal.uchicago.edu/dictionaries/biswas-bengali/.

Blankenship, B., Ladefoged, P., Bhaskararao, P., & Chase, N. (1993). Phonetic structures of Khonoma Angami. *UCLA Working Papers in Phonetics, 84*, 127–141.

Blevins, J. (2004). *Evolutionary Phonology: The emergence of sound patterns.* Cambridge: Cambridge University Press.

Blevins, J. (2006). A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics, 32* (2), 117–166.

Bodding, P. O. (1929–1936). *A Santali dictionary.* Oslo: I Kommisjon Hos Jacob Dybwab (Norwegian Academy of Science and Letters).

Boehm, E. D. (1998). *A phonological reconstruction of Proto-Tharu.* MA thesis. University of Texas at Arlington.

Boersma, P., & Hayes, B. (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry, 32* (1), 45–86.

Borchers, D. (2008). *A grammar of Sunwar: Descriptive grammar, paradigms, texts and glossary.* Leiden & Boston: Brill.

Bradley, D. (2003). Lisu. In G. Thurgood, & R. J. LaPolla (Eds.), *The Sino-Tibetan languages* (pp. 222–235). London & New York: Routledge.

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook, 3*, 219–252.

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 6*, 201–251.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica, 49*, 155–180.

Brown, J. (2008). *Theoretical aspects of Gitksan phonology.* Doctoral dissertation. University of British Columbia.

Brown, J. (2010). *Gitksan phonotactics.* Munich: LINCOM Europa.

Bubenik, V. (2003). Prākrits and Apabhraṁśa. In G. Cardona, & D. Jain (Eds.), *The Indo-Aryan Languages* (pp. 204–249). London & New York: Routledge.

Burling, R. (1957). Lushai phonemics. *Indian Linguistics, 17*, 148–155.

Burling, R. (2003). Garo. In G. Thurgood, & R. J. LaPolla (Eds.), *The Sino-Tibetan Languages* (pp. 387–400). London & New York: Routledge.

Burling, R., & Phom, L. A. (1998). Phom phonology and word list. *Linguistics of the Tibeto-Burman Area, 21* (2), 13–42.

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

Burrow, T., & Bhattacharya, S. (1960). Comparative vocabulary of the Gondi dialects. *Journal of the Asiatic Society, 2*, 73–251.

Burrow, T., & Bhattacharya, S. (1962). Gadba supplement. *Indo-Iranian Journal, 6* (1), 45–51.

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

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

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

Cain, B., & Gair, J. W. (2000). *Dhivehi (Maldivian).* Muenchen: Lincom Europa.

Caldwell, R. (1856). *A comparative grammar of the Dravidian or South-Indian family of languages* (1st ed.). London: Harrison.

Caldwell, R. (1875). *A comparative grammar of the Dravidian or South-Indian family of languages* (2nd ed.). London: Trübner and Co.

Cardona, G. (2003). Sanskrit. In G. Cardona, & D. Jain, *The Indo-Aryan languages* (pp. 104–160). London & New York: Routledge.

Cardona, G., & Suthar, B. (2003). Gujarati. In G. Cardona, & D. Jain, *The Indo-Aryan languages* (pp. 659–697). London & New York: Routledge.

Caughley, R. C. (1969). *Chepang phonemic summary*. Tibeto-Burman Phonemic Summaries, 4. Kathmandu: Summer Institute of Linguistics and Tribhuvan University.

Caughley, R. C. (1970). Chepang segmental synopsis. In A. Hale, & K. L. Pike (Eds.), *Tone systems of Tibeto-Burman languages of Nepal 1: Studies on tone and phonological segments* (pp. 279–299). Urbana: University of Illinois Press.

Chandrasekhar, A. (1969). The phonemes of Garhwali. *Indian Linguistics, 31* (3), 80–85.

Chatterji, S. K. (1970 [1926]). *The origin and development of the Bengali language*. London: George Allen & Unwin.

Chelliah, S. L. (1997). *A grammar of Meithei*. Berlin & New York: Mouton de Gruyter.

Chelliah, S. L. (2003). Meithei. In G. Thurgood, & R. J. LaPolla (Eds.), *The Sino-Tibetan languages* (pp. 427–438). London & New York: Routledge.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Christdas, P. (1988). *The phonology and morphology of Tamil*. Doctoral dissertation. Ithaca, NY: Cornell University.

Clements, G. N. (1980). *Vowel harmony in nonlinear generative phonology: An autosegmental model*. Indiana University Linguistics Club.

Clements, G. N. (1999). Affricates as noncontoured stops. In O. Fujimura, B. Joseph, & B. Palek (Eds.), *Item order in language and speech* (pp. 271–299). Prague: Charles University Press.

Clements, G. N. (2001). Representational economy in constraint-based phonology. In T. A. Hall (Ed.), *Distinctive feature theory* (pp. 71–146). Berlin: Mouton de Gruyter.

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

Cook, W. A. (1965). *A descriptive analysis of Mundari.* Doctoral dissertation. Washington, DC: Georgetown University.

Cooper, K. (1999). Dhimal. In Y. P. Yadava, & W. W. Glover (Eds.), *Topics in Nepalese linguistics* (pp. 26–44). Kathmandu: Royal Nepal Academy.

Coupe, A. R. (2007). *A grammar of Mongsen Ao.* Berlin & New York: Mouton de Gruyter.

Daniel, S., & Stephan, L. (2003). *A write-up on Paniya phonology*. Unpublished manuscript. Paniya Urgent Literacy for Social Empowerment (PULSE).

Dart, S. N., & Nihalani, P. (1999). The articulation of Malayalam coronal stops and nasals. *Journal of the International Phonetic Association, 29* (2), 129-142.

Das, A. R. (1977). *A study on the Nicobarese language.* Calcutta: Anthropological Survey of India, Government of India.

Das, A. S. (1973). *Structure of Malto.* Annamalainagar: Annamalai University.

Dasgupta, P. (2003). Bangla. In G. Cardona, & D. Jain, *The Indo-Aryan languages* (pp. 351–390). London & New York: Routledge.

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

DeLancey, S. (2003a). Classical Tibetan. In G. Thurgood, & R. J. LaPolla (Eds.), *The Sino-Tibetan Languages* (pp. 255–269). London & New York: Routledge.

DeLancey, S. (2003b). Lhasa Tibetan. In G. Thurgood, & R. J. LaPolla (Eds.), *The Sino-Tibetan Languages* (pp. 270–288). London & New York: Routledge.

Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review, 104* (1), 123–147.

Denwood, P. (1999). *Tibetan.* Amsterdam & Philadelphia: John Benjamins Publishing Company.

Deshpande, M. M. (1979). Genesis of Rgvedic retroflexion: a historical and sociolinguistic investigation. In M. M. Deshpande, & P. E. Hook, *Aryan and non-Aryan in India* (pp. 235–315). Ann Arbor: Centre for South and Southeast Asian Studies, University of Michigan.

Deshpande, M. M. (1992). Sanskrit. In W. Bright (Ed.), *International Encyclopedia of Linguistics* (Vol. 3, pp. 366–372). New York & Oxford: Oxford University Press.

Dev, A. (2009). Effect of retroflex sounds on the recognition of Hindi voiced and unvoiced stops. *AI & Society , 23*, 603-612.

Devaneyan, G. (1966). *The primary classical language of the world.* Ka[t]padi Extension, North Arcot District, Madras: Nesamani Publishing House.

Diffloth, G. F. (1975). The South Dravidian obstruent system in Irula. In H. F. Schiffman, & C. M. Eastman, *Dravidian phonological systems* (pp. 47–56). Seattle: Institute for Comparative and Foreign Area Studies, University of Washington.

Dixit, R. P., & Flege, J. E. (1991). Vowel context, rate and loudness effects on linguopalatal contact patterns in Hindi retroflex /ʈ/. *Journal of Phonetics, 19*, 213-229.

Dixon, R. M. (1980). *The languages of Australia.* Cambridge: Cambridge University Press.

Dixon, R. M. (2002). *Australian languages: Their nature and development.* Cambridge: Cambridge University Press.

Donegan, P. J., & Stampe, D. (2004). Munda lexical archive. Retrieved August 24, 2010 from: http://www.ling.hawaii.edu/austroasiatic/.

Dresher, B. E. (2003). The contrastive hierarchy in phonology. *Toronto Working Papers in Linguistics*, 47–62.

Dresher, B. E. (2009). *The contrastive hierarchy in phonology.* Cambridge: Cambridge University Press.

Dyszak, A. (1997). Upodobnienia głosek w ciągach fonicznych [Assimilations of segments in connected speech]. In A. Dyszak, E. Laskowska, & M. Żak-Święcicka (Eds.), *Fonetyczny i fonologiczny opis współczesnej polszczyzny [Phonetic and phonological description of contemporary Polish].* Bydgoszcz: Wydawnicwo Uczelni Wyższej Szkoły Pedagogicznej w Bydgoszczy.

Ebert, K. (1996). *Kodava.* München: Lincom Europa.

Ebert, K. (1997a). *A grammar of Athpare.* München: Lincom Eurpoa.

Ebert, K. (1997b). *Camling (Chamling).* München: Lincom Europa.

Ebert, K. (2003). Camling. In G. Thurgood, & R. J. LaPolla (Eds.), *The Sino-Tibetan languages* (pp. 533–545). London & New York: Routledge.

Efere, E. (2001). The pitch system of the Bumo dialect of Izon. (S. Gessner, S. Oh, & K. Shiobara, Eds.) *Current Research on African Languages and Linguistics (UBC Working Papers in Linguistics 4)*, 115-259.

Elfenbein, J. (1997a). Balochi phonology. In A. S. Kaye (Ed.), *Phonologies of Asia and Africa* (Vol. 2, pp. 761–776). Winona Lake, Indiana: Eisenbrauns.

Elfenbein, J. (1997b). Brahui phonology. In A. S. Kaye, *Phonologies of Asia and Africa* (Vol. 2, pp. 797–811). Winona Lake, Indiana: Eisenbrauns.

Elfenbein, J. (1997c). Pashto phonology. In A. S. Kaye (Ed.), *Phonologies of Asia and Africa* (Vol. 2, pp. 733–760). Winona Lake, Indiana: Eisenbrauns.

Elfenbein, J. (1998). Brahui. In S. B. Steever, *The Dravidian languages* (pp. 388–414). London & New York: Routledge.

Emeneau, M. B. (1939). The vowels of the Badaga language. *Language, 15* (1), 43–47.

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

Emeneau, M. B. (1961). *Kolami: A Dravidian language.* Annamalainagar: Annamalai University.

Ernest, & Ernest, S. (2000). *A summary of Kurumba Kannada phonology*. Unpublished manuscript. Bangalore: Carey Bible Translators.

Evans, N. (1995). Current issues in Australian languages. In J. A. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 723–61). Oxford: Blackwell Publishers.

Fabricius, J. P. (1972). *J. P. Fabricius's Tamil and English dictionary* (4th ed.). Tranquebar: Evangelical Lutheran Mission Pub. House. Online URL: http://dsal.uchicago.edu/dictionaries/fabricius/.

Fernandez, F. (1968). *A grammatical sketch of Remo: A Munda language*. Doctoral dissertation. University of North Carolina.

Flemming, E. (1995). *Auditory representations in phonology*. Doctoral dissertation. Los Angeles: University of California.

Flemming, E. (2004). Contrast and perceptual distinctiveness. In B. Hayes, R. Kirchner, & D. Steriade, *Phonetically based phonology* (pp. 232-276). Cambridge: Cambridge University Press.

Frisch, S. (1996). *Similarity and frequency in phonology.* Doctoral dissertation. Northwestern University.

Frisch, S., Pierrehumbert, J., & Broe, M. (2004). Similarity avoidance and the OCP. *Natural Language and Linguistic Theory, 22*, 179–228.

Gafos, A. (1999). *The articulatory basis of locality in phonology.* New York & London: Garland Publishing.

Gair, J. W. (2003). Sinhala. In G. Cardona, & D. Jain, *The Indo-Aryan languages* (pp. 766–817). London & New York: Routledge.

Gallagher, G. (2010). Perceptual distinctness and long-distance laryngeal restrictions. *Phonology, 27*, 435–480.

Gallagher, G. (2012). Perceptual similarity in non-local laryngeal restrictions. *Lingua, 122*, 112–124.

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

Genetti, C. (2003). Dolakhā Newār. In G. Thurgood, & R. J. LaPolla (Eds.), *The Sino-Tibetan languages* (pp. 355–370). London & New York: Routledge.

Genetti, C. (2007). *A grammar of Dolakha Newar.* Berlin & New York: Mouton de Gruyter.

George, B. J., & Joseph, C. (2008). *A write-up on Korwa phonology.* Unpublished manuscript.

Ghai, V. K. (1991). *Studies in phonetics and phonology (with special reference to Dogri).* New Delhi: Ariana Publishing House.

Ghosh, A. (2008). Santali. In G. D. Anderson (Ed.), *The Munda languages* (pp. 11–98). London & New York: Routledge.

Giridhar, P. P. (1994). *Mao Naga grammar.* Mysore: Central Institute of Indian Languages.

Glover, W. W. (1969). *Gurung phonemic summary.* Kirtipur: Summer Institute of Linguistics and Tribhuvan University.

Glover, W. W. (1970). Gurung segmental synopsis. In A. Hale, & K. L. Pike (Eds.), *Tone systems of Tibeto-Burman languages of Nepal 1: Studies on tone and phonological segments* (pp. 211–236). Urbana: University of Illinois Press.

Gnanadesikan, A. E. (1994). The geometry of coronal articulations. In M. Conzalez (Ed.), *NELS 24: Proceedings of the North East Linguistic Society* (pp. 125–139). Amherst: University of Massechusetts.

Gnanasekaran, & Sheeba. (2003). *Rana Tharu phonemic summary*. Unpublished manuscript. Indian Evangelical Mission.

Goldsmith, J. A. (1979). *Autosegmental phonology*. New York: Garland.

Gordon, K., & Schoettelndreyer, B. (1970). Sherpa segmental synopsis. In A. Hale, & K. L. Pike (Eds.), *Tone systems of Tibeto-Burman languages of Nepal 1: Studies on tone and phonological segments* (pp. 345-67). Urbana: University of Illinois Press.

Goswami, G. C. (1966). *An introduction to Assamese phonology*. Poona: Deccan College Postgraduate and Research Institute.

Goswami, G. C., & Tamuli, J. (2003). Asamiya. In G. Cardona, & D. Jain, *The Indo-Aryan languages* (pp. 391–443). London & New York: Routledge.

Goswami, K. K. (2000). *Panjabi-English English-Panjabi dictionary*. New Delhi: UBS Publishers's Distribuors.

Goswami, U. (1994). *An introduction to the Deuri language*. Guwahati, Assam: Anundoram Borooah Institute of Language, Art and Culture.

Gowda, K. S. (1972). *Ao-Naga phonetic reader*. Mysore: Central Institute of Indian Languages.

Griffiths, A. (2008). Gutob. In G. D. Anderson (Ed.), *The Munda languages* (pp. 633–681). London & New York: Routledge.

Grignard, A. (1986 [1924]). *An Oraon-English dictionary.* New Delhi: Unity Book Service.

Gumperz, J. J. (1957). Notes on the phonology of Mundari. *Indian Linguistics, 17*, 6–15.

Gupta, K. D. (1963). *An introduction to the Gallong language.* Shillong: North-East Frontier Agency.

Gusain, L. (2000). *Bagri.* Muenchen: Lincom Europa.

Gusain, L. (2001). *Shekhawati.* Muenchen: Lincom Europa.

Gusain, L. (2003). *Mewati.* Muenchen: Lincom Europa.

Gusain, L. (2004). *Marwari.* Muenchen: Lincom Europa.

Gwynn, J. P. (1991). *A Telugu-English dictionary.* Delhi and New York: Oxford University Press. Online URL: http://dsal.uchicago.edu/dictionaries/gwynn/.

Höhlig, M., & Hari, M. (1976). *Kagate phonemic summary.* Kathmandu: Summer Institute of Linguistics and Institute of Nepal and Asian Studies.

Hahn, F. (1911). *Kurukh grammar.* Calcutta: Bengal Secretariat Press.

Hale, A. (1970). Newari segmental synopsis. In A. Hale, & K. L. Pike (Eds.), *Tone systems of Tibeto-Burman languages of Nepal 1: Studies on tone and phonological segments* (pp. 300–327). Urbana: University of Illinois Press.

Hall, T. A. (1997a). The historical development of retroflex consonants in Indo-Aryan. *Lingua, 102* (4), 203-221.

Hall, T. A. (1997b). *The phonology of coronals.* Amsterdam and Philadelphia: John Benjamins Publishing Company.

Hall, T. A. (2007). Segmental features. In P. de Lacy (Ed.), *The Cambridge handbook of phonology* (pp. 311–334). Cambridge University Press.

Hallberg, D. G., & Hallberg, C. E. (1999). *Indus Kohistani: A preliminary phonological and morphological analysis.* Islamabad: National Institute of Pakistan Studies, Quaid-i-Azam University and Summer Institute of Linguistics.

Halle, M., & Vergnaud, J.-R. (1981). Harmony processes. In W. Klein, & W. J. Levelt (Eds.), *Crossing the boundaries in linguistics* (pp. 1-22). Dordrecht: Reidel.

Hamann, S. (2002). Retroflexion and retraction revised. *ZAS Papers in Linguistics, 28*, 13-26.

Hamann, S. (2003). *The phonetics and phonology of retroflexes.* Utrecht: LOT (Landelijke Onderzoekschool Taalwetenschap).

Hamann, S. (2005). The diachronic emergence of retroflex segments in three languages. *Link, 15* (1), 29-48.

Hamann, S., & Fuchs, S. (2010). Retroflexion of voiced stops: data from Dhao, Thulung, Afar and German. *Language and Speech, 53* (2), 181-216.

Hamilton, P. (1996). *Phonetic constraints and markedness in the phonotactics of Australian aboriginal languages.* Doctoral dissertation. University of Toronto.

Hamp, E. P. (1996). On the Indo-European origins of the retroflexes in Sanskrit. *Journal of the American Oriental Society, 116* (4), 719–723.

Handoo, J. (1973). *Kashmiri phonetic reader.* Mysore: Central Institute of Indian Languages.

Hansson, G. Ó. (2001). *Theoretical and typological issues in consonant harmony*. Doctoral dissertation. University of California, Berkeley.

Hansson, G. Ó. (2007). Blocking effects in agreement by correspondence. *Linguistic Inquiry, 38* (2), 395-409.

Hansson, G. Ó. (2010). *Consonant harmony: Long-distance interaction in phonology.* Berkeley: University of California Press.

Hardwick, M. (1984). *Tahltan phonology and morphology*. MA thesis. University of Toronto.

Hargreaves, D. (2003). Kathmandu Newar (Nepāl Bhāśā). In G. Thurgood, & R. J. LaPolla (Eds.), *The Sino-Tibetan Languages* (pp. 371–384). London & New York: Routledge.

Hari, M. (1969). *Thakali phonemic summary.* Kathmandu: Summer Institute of Linguistics and Tribhuvan University.

Hari, M. (1970). Thakali segmental synopsis. In A. Hale, & K. L. Pike (Eds.), *Tone systems of Tibeto-Burman languages of Nepal 1: Studies on tone and phonological segments* (pp. 258–78). Urbana: University of Illinois Press.

Heegård, J., & Mørch, I. E. (2004). Retroflex vowels and other peculiarities in the Kalasha sound system. In A. Saxena (Ed.), *Himalayan languages, past and present* (pp. 57–76). Berlin & New York: Mouton de Gruyter.

Hildebrandt, K. A. (2004). A grammar and glossary of the Manange language. In C. Genetti (Ed.), *Tibeto-Burman languages of Nepal: Manange and Sherpa* (pp. 1–189). Canberra: Research School of Pacific and Asian Studies, The Australian National University.

Hock, H. H. (1991). *Principles of historical linguistics.* (2nd, Ed.) Berlin & New York: Mouton de Gruyter.

Hockings, P., & Pilot-Raichoor, C. (1992). *A Badaga-English Dictionary.* Berlin & New York: Mouton de Gruyter.

Honda, I. (2003). A sketch of Tangbe. In T. R. Kansakar, & M. Turin (Eds.), *Themes in Himalayan languages and linguistics* (pp. 49–64). Heidelberg and Kathmandu: South Asia Institute and Tribhuvan University.

Huber, B. (2005). *The Tibetan dialect of Lende (Kyirong): a grammatical description with historical annotations.* Bonn: VGH Wissenschaftsverlag GmbH.

Hyslop, G. (2008). Kurtöp phonology in the context of North East India. In S. Morey, & M. Post (Eds.), *North East Indian linguistics* (pp. 3–25). New Delhi: Cambridge University Press India Pvt. Ltd.

Immanuel, & Jane. (2003). *A write-up on Pawri phonology*. Unpublished manuscript. FMPB.

Israel, M. (1979). *A grammar of the Kuvi language (with texts and vocabulary).* Trivandrum: Dravidian Linguistics Association.

Jacquesson, F. (2005). *Le Deuri: langue Tibéto-Birmane d'Assam.* Leuven: Peeters.

Jagannath. (1981). *Telugu loanword phonology.* Doctoral dissertation. University of Arizona.

Jain, B. D. (1934). *A phonology of Panjābī as spoken about Ludhiana and a Ludhiānī phonetic reader.* Lahore: The University of the Panjab.

Jaiswal, M. P. (1962). *A linguistic study of Bundeli.* Leiden: E. J. Brill.

Jarmasz, L.-G. (2008). *Complexity of structure and distinctness in the polish coronal inventory.* Unpublished generals paper. University of Toronto.

Jeyapaul, V. Y. (1987). *Karbi grammar.* Manasagangotri, Mysore: Central Institute for Indian Languages.

Jha, S. K. (2001). *Maithili: Some aspects of its phonetics and phonology.* Delhi: Motilal Banarsidass.

Jordan-Horstmann, M. (1969). *Sadani: A Bhojpuri dialect spoken in Chotanagpur.* Wiesbaden: Otto Harrassowitz.

Joseph, U. V. (2007). *Rabha.* Leiden & Boston: Brill.

Jun, J. (2004). Place assimilation. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology* (pp. 58-86). Cambridge: Cambridge University Press.

Jurgec, P. (2010). *Feature spreading 2.0: A unified theory of assimilation.* Doctoral dissertation. University of Tromsø.

Karapurkar, P. (1972). *Tripuri phonetic reader.* Mysore: Central Institute of Indian Languages.

Karapurkar, P. (1976). *Kokborok grammar.* Mysore: Central Institute of Indian Languages.

Katre, S. M. (1966). *The formation of Konkani.* Poona: Deccan College Postgraduate and Research Institute.

Kawahara, S., Ono, H., & Sudo, K. (2006). Consonant cooccurrence restrictions in Yamoto Japanese. *Japanese and Korean Linguistics, 14*, 27-38.

Kaye, A. S. (1997). Hindi-Urdu phonology. In A. S. Kaye, *Phonologies of Asia and Africa* (Vol. 2, pp. 637–652). Winona Lake, Indiana: Eisenbrauns.

Keane, E. (2004). Tamil. *Journal of the International Phonetic Association, 34* (1), 111–116.

Keating, P. A. (1991). Coronal places of articulation. In C. Paradis, & J.-F. Prunet (Eds.), *The special status of coronals: Internal and external evidence* (pp. 29–48). San Diego: Academic Press.

Kelly, B. (2004). A grammar and glossary of the Sherpa language. In C. Genetti (Ed.), *Tibeto-Burman languages of Nepal: Manange and Sherpa* (pp. 193–324). Canberra: Research School of Pacific and Asian Studies, The Australian National University.

Kelmakov, V. K. (1987). Prapermskie affricaty v sovremennyx udmurtskix dialektax [Proto-Permic affricates in modern Udmurt dialects]. *Sovetskoe finno-ugrovedenie [Soviet Finno-Ugric Studies], 23* (1), 34–44.

Khatiwada, R. (2007). Nepalese retroflex stops: a static palatogrphy study of inter- and intra-speaker variability. Paper presented at Interspeech 2007, Antwerp.

Khatiwada, R. (2009). Nepali. *Journal of the International Phonetic Association, 39* (3), 373-380.

Khubchandani, L. M. (2003). Sindhi. In G. Cardona, & D. Jain, *The Indo-Aryan languages* (pp. 622–658). London & New York: Routledge.

King, J. T. (2008). *A grammar of Dhimal.* Doctoral dissertation. Leiden University.

Kirivasan, & Amirthamary. (2000). *Bhatri phonemic summary.* Unpublished manuscript. Indian Evangelical Mission.

Kobayashi, M., & Murmu, G. (2008). Keraʔ Mundari. In G. D. Anderson (Ed.), *The Munda languages* (pp. 165–194). London & New York: Routledge.

Kochetov, A. (2007). *Phonetics and phonology of Komi-Permyak coronals: Root co-occurrence restrictions.* Paper presented at the Coronal Workshop, University of Toronto, November 9, 2007.

Koshal, S. (1976). *Ladakhi phonetic reader.* Mysore: Central Institute of Indian Languages.

Koshal, S. (1978). *Conflicting phonological patterns: A study in the adaptation of English loan words in Hindi.* New Delhi: Bahri Publications.

Koshal, S. (1979). *Ladakhi grammar.* Delhi: Motilal Banarsidass.

Kotapish, C., & Kotapish, S. (1973). *Darai phonemic summary.* Kathmandu: Summer Institute of Linguistics and Institute of Nepal and Asian Studies.

Koul, O. N. (2003). Kashmiri. In G. Cardona, & D. Jain (Eds.), *The Indo-Aryan languages* (pp. 895–952). London & New York: Routledge.

Krishan, S. (1980). *Thadou: a grammatical sketch.* Calcutta: Anthropological Survey of India, Government of India.

Krishnamurti, B. (1969). *Konda or Kubi: A Dravidian language.* Hyderabad: Tribal Cultural Research & Training Institute, Government of Andhra Pradesh.

Krishnamurti, B. (1998). Telugu. In S. B. Steever, *The Dravidian languages* (pp. 202–240). London & New York: Routledge.

Krishnamurti, B. (2003). *The Dravidian languages.* Cambridge: Cambridge University Press.

Krishnamurti, B., & Benham, B. A. (1998). Konda. In S. B. Steever, *The Dravidian languages* (pp. 241–269). London & New York: Routledge.

Kristoffersen, G. (2000). *The phonology of Norwegian.* Oxford: Oxford University Press.

Kuegler, K. P., & Kuegler, D. (1974). *Danuwar Rai phonemic summary.* Kathmandu: Summer Institute of Linguistics and Institute of Nepal and Asian Studies, Tribhuvan University.

Kuipers, A. H. (1974). *The Shuswap language: grammar, texts, dictionary.* The Hague & Paris : Mouton.

Kulkarni, S. B. (1976). *Bhili of Dangs.* Poona: Deccan College.

Kumari, B. S. (1972). *Malayalam phonetic reader.* Mysore: Central Institute of Indian Languages.

Kuriakkose, L., & Liju, L. (2008). *A write-up on Kodaku phonology.* Unpublished manuscript.

Kurian, B., & Kurian, M. (2000). *Phonemic analysis of Southern Gondi.* Unpublished manuscript.

Kuroda, S. -Y. (1967). *Yawelmani phonology.* Cambridge, MA: MIT Press.

Ladefoged, P., & Bhaskararao, P. (1983). Non-quantal aspects of consonant production: A study of retroflex consonants. *Journal of Phonetics, 11*, 291–302.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages.* Oxford: Blackwell Publishers.

Lahaussois, A. (2002). *Aspects of the grammar of Thulung Rai: An endangered Himalayan language*. Doctoral dissertation. Berkeley: University of California.

Lahaussois, A. (2003). Thulung Rai. *Himalayan Linguistics, 1*, 1–25.

Lahaussois, A. (2009). Koyi Rai: An initial grammatical sketch. *Himalayan Linguistics, 4*, 1–33.

Lal, S. M. (1991). *A descriptive analysis of Urali.* Mysore: Central Institute of Indian Languages.

Lehmann, T. (1998). Old Tamil. In S. B. Steever, *The Dravidian languages* (pp. 75–99). London & New York: Routledge.

Levitt, S. H. (2010). Fortunatov's law and Dravidian. *International Journal of Dravidian Linguistics, 39* (1), 19–90.

Lewis, M. P. (Ed.). (2009). *Ethnologue: Languages of the world* (Sixteenth ed.). Dallas: SIL International.

Liljegren, H. (2008). *Towards a grammatical description of Palula: An Indo-Aryan language of the Hindu Kush*. Doctoral dissertation. Stockholm University.

Liljegren, H., & Haider, N. (2009). Palula. *Journal of the International Phonetic Association, 39* (3), 381–386.

Lincoln, N. J. (1969). *A descriptive analysis of the Adilabad dialect of Gondi.* Doctoral dissertation. Cornell University.

Lobsang, G. H. (1995). *A short sketch of Balti grammar: a Tibetan dialect spoken in northern Pakistan.* Berne, Switzerland: Institut für Sprachwissenschaft, Universität Bern.

Lorimer, D. L. (1935). *The Burushaski language. Volume I: Introduction and grammar.* Oslo: Instituttet for Sammenlignende Kulturforskning.

Lorimer, D. L. (1938). *The Burushaski language. Volume III: Vocabularies and index.* Oslo: Instituttet for Sammenlignende Kulturforskning.

Lorimer, D. L. (1939). *The Dumāki language.* Nijmegan: Dekker and Van de Vegt N V.

Lunsford, W. A. (2001). *An overview of linguistic structures in Torwali, a language of northern Pakistan*. MA thesis. University of Texas at Arlington.

Mørch, I. E., & Heegård, J. (1997). *Retroflekse vokalers oprindelse i kalashamon i historisk og areallingvistisk perspektiv* [The Origin of Retroflex Vowels in Kalashamon in a Historical and Areal Linguistic Perspective]. MA thesis. Department of General and Applied Linguistics, University of Copenhagen.

MacEachern, M. R. (1997). *Laryngeal cooccurrence restrictions*. Doctoral dissertation. Los Angeles: University of California.

Mackenzie, S. (2005). Similarity and contrast in consonant harmony systems. *Toronto Working Papers in Linguistics*, 169–182.

Mackenzie, S. (2009). *Contrast and similarity in consonant harmony processes.* Doctoral dissertation. University of Toronto.

Mackenzie, S. (2011). Contrast and the evaluation of similarity: Evidence from consonant harmony. *Lingua*, *121*, 1401–1423.

Maddieson, I. (1984). *Patterns of sounds.* Cambridge: Cambridge University Press.

Maffei, A. F. (1883). *A Konkani-English dictionary.* Mangalore: Basel Mission Press.

Mahanta, S. (2007). *Directionality and locality in vowel harmony: With special reference to vowel harmony in Assamese.* Utrecht: LOT (Landelijke Onderzoekschool Taalwetenschap).

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, A. N. (1995). *The phonology and morphology of Panjabi.* New Delhi: Munshiram Manoharlal Publishers.

Mallikarjun, B. (1993). *A descriptive analysis of Yerava.* Mysore: Central Institute of Indian Languages.

Manoharan, S. (1989). *A descriptive and comparative study of Andamanese language.* Calcutta: Anthropological Survey of India, Government of India.

Martin, A. T. (2005). *The effects of distance on lexical bias: Sibilant harmony in Navajo compounds.* MA Thesis. University of California, Los Angeles.

Masica, C. P. (1976). *Definining a linguistic area: South Asia.* Chicago & London: University of Chicago Press.

Masica, C. P. (1991). *The Indo-Aryan languages.* Cambridge: Cambridge University Press.

Masica, C. P. (1992). South Asian languages. In W. Bright (Ed.), *International Encyclopedia of Linguistics* (Vol. 4, pp. 38–42). New York & Oxford: Oxford University Press.

Mathew, S., & Mathew, A. (2003). *A summary of Juang phonology*. Unpublished manuscript.

Mathew, S., & Susan, M. (2000). *Dungra Bhil phonemic summary*. Unpublished manuscript. FMPB.

Mathews, S. (2003). *A summary of Desiya phonology*. Unpublished manuscript. Koraput District, Orissa: Asha Kiran Society.

Matisoff, J. A. (2003). *Handbook of Proto-Tibeto-Burman: System and philosophy of Sino-Tibetan reconstruction.* Berkeley: University of California Press.

Matson, D. M. (1964). *A grammatical sketch of Juang, a Munda language.* Doctoral dissertation. University of Wisconsin.

Mazaudon, M. (2003). Tamang. In G. Thurgood, & R. J. LaPolla (Eds.), *The Sino-Tibetan Languages* (pp. 291–314). London & New York: Routledge.

Mazaudon, M., & Michailovsky, B. (1988). Lost syllables and tone contour in Dzongkha (Bhutan). In D. Bradley, E. J. Henderson, & M. Mazaudon (Eds.), *Prosodic analysis and Asian linguistics: to honour R. K. Sprigg* (pp. 115–136). Pacific Linguistics, C–104.

McCarthy, J. J. (1986). OCP effects: Gemination and antigemination. *Linguistic Inquiry, 17*, 207-263.

McCarthy, J. J. (1988). Feature geometry and dependency: a review. *Phonetica, 43*, 84-108.

McCarthy, J. J. (1994). The phonetics and phonology of Semitic pharyngeals. In P. Keating (Ed.), *Papers in laboratory phonology III* (pp. 191-283). Cambridge: Cambridge University Press.

McCarthy, J. J. (2002). On targeted constraints and cluster simplification. *Phonology, 19* (2), 273-292.

McGregor, R. S. (1993). *The Oxford Hindi-English dictionary.* Oxford: Oxford University Press.

Michailovsky, B. (2003). Hayu. In G. Thurgood, & R. J. LaPolla (Eds.), *The Sino-Tibetan languages* (pp. 518–532). London & New York: Routledge.

Miranda, R. V. (2003). Konkani. In G. Cardona, & D. Jain, *The Indo-Aryan languages* (pp. 729–765). London & New York: Routledge.

Misra, B. G. (1967). *Historical phonology of Modern Standard Hindi: Proto-Indo-European to the present.* Doctoral dissertation. Cornell University.

Misra, P. S. (1986). Phonemes of the Dimasa language. *Indian Linguistics, 47*, 33–37.

Mistry, P. J. (1997). Gujarati phonology. In A. S. Kaye, *Phonologies of Asia and Africa* (Vol. 2, pp. 653–673). Winona Lake, Indiana: Eisenbrauns.

Molesworth, J. T. (1857). *A dictionary, Marathi and English* (Second ed.). Bombay: Printed for government at the Bombay Education Society's press. Online URL: http://dsal.uchicago.edu/dictionaries/molesworth/.

Morey, S. (2005). *The Tai languages of Assam – a grammar and texts.* Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Morgenstierne, G. (1973). *Indo-Iranian frontier languages, vol. 4, the Kalasha language.* Oslo: Universitetsforlaget.

Munshi, S. (2006). *Jammu and Kashmir Burushaski: Language, language contact, and change.* Doctoral dissertation. University of Texas at Austin.

Nagaraja, K. S. (1985). *Khasi: A descriptive analysis.* Pune: Deccan College Post-Graduate and Research Institute.

Nagaraja, K. S. (1989). *Khasi phonetic reader.* Mysore: Central Institute of Indian Languages.

Natarajan, G. V. (1985). *Abujhmaria grammar.* Mysore: Central Institute of Indian Languages.

Neukom, L. (2001). *Santali.* München: Lincom Europa.

Ní Chiosáin, M., & Padgett, J. (1997). *Markedness, segment realisation, and locality in spreading. Report no. LRC-97-01.* Santa Cruz, CA: Linguistics Research Center, UCSC.

Ní Chiosáin, M., & Padgett, J. (2001). Markedness, segment realization, and locality in spreading. In L. Lombardi (Ed.), *Segmental phonology in Optimality Theory: Constraints and representations* (pp. 118–156). Cambridge: Cambridge University Press.

Nihalani, P. (1999). Sindhi. In *Handbook of the International Phonetic Association* (pp. 131–134). Cambridge: Cambridge University Press.

Noonan, M. (2003a). Chantyal. In G. Thurgood, & R. J. LaPolla (Eds.), *The Sino-Tibetan Languages* (pp. 315–335). London & New York: Routledge.

Noonan, M. (2003b). Nar Phu. In G. Thurgood, & R. J. LaPolla (Eds.), *The Sino-Tibetan Languages* (pp. 336–352). London & New York: Routledge.

Norihiko, U. (1991). *The language of the Saurashtrans in Tirupati.* Bangalore: Mahalaxmi Enterprises.

Oberlies, T. (2003). Aśokan Prakrit and Pāli. In G. Cardona, & D. Jain (Eds.), *The Indo-Aryan languages* (pp. 161–203). London & New York: Routledge.

Ohala, M. (1978). Conflicting expectations for the direction of sound change. *Indian Linguistics, 39*, 25–28.

Ohala, M. (1983). *Aspects of Hindi phonology.* Delhi: Motilal Banarsidass.

Ohala, M. (1994). Hindi. *Journal of the International Phonetic Association, 24* (1), 35–38.

Opgenort, J. R. (2004). *A grammar of Wambule: grammar, lexicon, texts, and cultural survey of a Kiranti tribe of eastern Nepal.* Leiden & Boston: Brill.

Opgenort, J. R. (2005). *A grammar of Jero: with a historical comparative study of the Kiranti languages.* Leiden & Boston: Brill.

Osada, T. (2008). Mundari. In G. D. Anderson (Ed.), *The Munda languages* (pp. 99–164). London & New York: Routledge.

Pali Text Society. (1921–1925). *The Pali Text Society's Pali-English dictionary.* London: Pali Text Society. Online URL: http://dsal.uchicago.edu/dictionaries/pali/.

Pandharipande, R. (1997). *Marathi.* London & New York: Routledge.

Pandharipande, R. (2003). Marathi. In G. Cardona, & D. Jain, *The Indo-Aryan languages* (pp. 698–728). London & New York: Routledge.

Patel, V. (1999). *A preliminary summary of Garasia phonology*. Unpublished manuscript.

Patnaik, M. (2008). Juang. In G. D. Anderson (Ed.), *The Munda languages* (pp. 508–556). London & New York: Routledge.

Penny, M., Penny, J., & Rao, P. D. (2005). *Gondi dictionary.* Utnoor, Andhra Pradesh: Integrated Tribal Development Agency.

Periyalwar, R. (1979). *Phonology of Irula with vocabulary.* Annamalainagar: Annamalai University.

Peterson, D. A. (2003). Hakha Lai. In G. Thurgood, & R. J. LaPolla (Eds.), *The Sino-Tibetan languages* (pp. 409–426). London & New York: Routledge.

Peterson, J. (2008). Kharia. In G. D. Anderson (Ed.), *The Munda languages* (pp. 434–507). London & New York: Routledge.

Peterson, J. (2009). A Kharia-English lexicon. *Himalayan Linguistics Archive 5*, i-xv, 1-212.

Pfeiffer, M. (1972). *Elements of Kurux historical phonology.* Leiden: E. J. Brill.

Pierrehumbert, J. (1993). Dissimilarity in the Arabic verbal roots. *Proceedings of the North East Linguistics Society*, *23*, pp. 367-381.

Piggott, G. L. (1992). Variability in feature dependency: The case of nasality. *Natural Language & Linguistic Theory, 10* (1), 33–77.

Pillai, V. C. (1976). *Phonology of Kasaba with vocabulary.* Annamalainagar: Annamalai University.

Pinnow, H.-J. (1964). Bemerkungen zur Phonetik und Phonemik des Kurukh. *Indo-Iranian Journal, 8* (1), 32-59.

Plaisier, H. (2003). Lepcha. In G. Thurgood, & R. J. LaPolla (Eds.), *The Sino-Tibetan languages* (pp. 705–716). London & New York: Routledge.

Plaisier, H. (2007). *A grammar of Lepcha.* Leiden & Boston: Brill.

Platts, J. T. (1884). *A dictionary of Urdu, classical Hindi, and English.* London: W. H. Allen & Co. Online URL: http://dsal.uchicago.edu/dictionaries/platts/.

Poser, W. J. (1982). Phonological representations and action-at-a-distance. In H. van der Hulst, & N. Smith (Eds.), *The structure of phonological representations, Part II* (pp. 121–158). Dordrecht: Foris Publications.

Poudel, K. P. (2006). *Dhankute Tamang grammar.* Muenchen: Lincom Europa.

Pozdniakov, K., & Segerer, G. (2007). Similar place avoidance: A statistical universal. *Linguistic Typology, 11*, 307-348.

Prasad, B. R. (1991). *Mising grammar.* Mysore: Central Institute of Indian Languages.

Prince, A., & Smolensky, P. (2004 [1993]). *Optimality theory: Constraint interaction in generative grammar.* Oxford: Blackwell. [Originally published in 1993 as RuCCS Technical Report No. 2, Rutgers Center for Cognitive Science, Rutgers University].

Pruitt, J. S., Jenkins, J. J., & Strange, W. (2006). Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. *Journal of the Acoustical Society of America, 119* (3), 1684–1696.

Pruitt, J. S., Strange, W., Polka, L., & Aguilar, M. C. (1990). Effects of category knowledge and syllable truncation during auditory training on Americans' discrimination of Hindi retroflex-dental contrasts. *Journal of the Acoustical Society of America*, S72.

Qingxia, D., & Diehl, L. (2003). Jinghpo. In G. Thurgood, & R. J. LaPolla (Eds.), *The Sino-Tibetan languages* (pp. 401–408). London & New York: Routledge.

Radhakrishnan, R. (1981). *The Nancowry word, phonology, affixal morphology and roots of a Nicobarese language.* Edmonton, Alberta: Linguistic Research Inc.

Radloff, C. F. (1999). *Aspects of the sound system of Gilgiti Shina.* Islamabad: National Institute of Pakistan Studies, Quaid-i-Azam University and Summer Institute of Linguistics.

Rajan, F. H., & Rajan, J. (2001). *Gutob-Gadaba phonemic summary.* Lamtaput, Orissa: Asha Kiran Society.

Rajaram, S. (1972). *Tamil phonetic reader.* Mysore: Central Institute of Indian Languages.

Ramamurti, G. V. (1986 [1933]). *Sora-English dictionary.* Delhi: Mittal Publications.

Ramanujan, A. K., & Masica, C. (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* (pp. 543–577). Paris: Mouton.

Ramaswami, N. (1975). *Brokskat phonetic reader.* Mysore: Central Institute of Indian Languages.

Ramaswami, N. (1992). *Bhumij grammar.* Mysore: Central Institute of Indian Languages.

Rangan, K. (1975). *Balti phonetic reader.* Mysore: Central Institute of Indian Languages.

Rao, V. S., & Patnaik, R. K. (1992). *Gadaba: The language and the people.* Amaravathi, Andhra Pradesh: Sri Papayaradhya Sahita Kendram.

Rapold, C. J. (2006). *Towards a grammar of Benchnon.* Doctoral dissertation. Leiden University.

Ravindran, N. (1974). *Angami phonetic reader.* Mysore: Central Institute of Indian Languages.

Ray, T. S. (2003). Oriya. In G. Cardona, & D. Jain, *The Indo-Aryan languages* (pp. 444–476). London & New York: Routledge.

Reddy, B. R., Upadhyaya, S. P., & Reddy, J. (1974). *Kuvi phonetic reader.* Mysore: Central Institute of Indian Languages.

Reddy, K. N. (1986). A palatographic study of Telugu retroflex consonants. In K. Rangan (Ed.), *Proceedings of the Thirteenth All India Conference of Dravidian Linguists* (pp. 163-186). Thanjavur: Tamil University.

Rehman, K. A., & Baart, J. L. (2005). A first look at the language of Kundal Shahi in Azad Kashmir. *SIL Electronic Working Papers 2005-008* (pp. 1–22). Retrieved November 28, 2008 from http://www.sil.org/silewp/2005/silewp2005-008.pdf.

Reynolds, C. (2003). *A Maldivian dictionary.* London & New York: RoutledgeCurzon.

Riccardi, T. (2003). Nepali. In G. Cardona, & D. Jain, *The Indo-Aryan languages* (pp. 538–580). London & New York: Routledge.

Rose, S. (2011). Long-distance assimilation of consonants. In v. Oostendorp, Marc, C. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell Companion to Phonology* (pp. 1811-1837). Oxford: Wiley-Blackwell.

Rose, S., & Walker, R. (2004). A typology of consonant agreement as correspondence. *Language, 80* (3), 475-531.

Sagey, E. C. (1986). *The representation of features and relations in non-linear phonology.* Doctoral dissertation. MIT.

Sahu, R. N. (1980). Notes on the phonology of Brijia. *Indian Linguistics, 41* (3–4), 157–160.

Saksena, B. (1937). *Evolution of Awadhi.* Allahabad: The Indian Press.

Sakthivel, S. (1976). *Phonology of Toda with vocabulary.* Annamalainagar: Annamalai University.

Sakthivel, S. (1977). *A grammar of the Toda language.* Annamalainagar: Annamalai University.

Sastry, G. D. (1984a). *Mishmi grammar.* Mysore: Central Institute of Indian Languages.

Sastry, G. D. (1984b). *Mishmi phonetic reader.* Mysore: Central Institute of Indian Languages.

Sastry, J. V. (1972). *Telugu phonetic reader.* Mysore: Central Institute of Indian Languages.

Satish, U. S. (1990). *A linguistic study of Jaunsari.* New Delhi: Creative.

Schein, B., & Steriade, D. (1986). On geminates. *Linguistic Inquiry, 17* (4), 691-744.

Schmidt, R. L. (2003). Urdu. In G. Cardona, & D. Jain, *The Indo-Aryan languages* (pp. 286–350). London & New York: Routledge.

Schmidt, R. L., & Kohistani, R. (2008). *A grammar of the Shina language of Indus Kohistan.* Wiesbaden: Harrassowitz Verlag.

Schwarzschild, L. A. (1973). Initial retroflex consonants in Middle Indo-Aryan. *Journal of the American Oriental Society, 93* (4), 482-487.

Sebeok, T. A. (1943). Phonemic system of Santali. *Journal of the American Oriental Society, 63* (1), 66–67.

Selvaraj, D., & Selvaraj, J. (2003). *Betta Kurumba phonemic summary.* Unpublished manuscript. New Life Computer Institute.

Shackle, C. (1976). *The Siraiki language of Central Pakistan.* London: School of Oriental and African Studies, University of London.

Shackle, C. (2003). Panjabi. In G. Cardona, & D. Jain, *The Indo-Aryan languages* (pp. 581–621). London & New York: Routledge.

Shalev, M., Ladefoged, P., & Bhaskararao, P. (1993). Phonetics of Toda. *UCLA Working Papers in Phonetics, 84*, 89–125.

Shapiro, M. C. (2003). Hindi. In G. Cardona, & D. Jain, *The Indo-Aryan languages* (pp. 250–285). London & New York: Routledge.

Sharma, D. D. (1988). *A descriptive grammar of Kinnauri.* Delhi: Mittal Publications.

Sharma, J. C. (1979). *Gojri phonetic reader.* Mysore: Central Institute of Indian Languages.

Sharma, S. (1974). *Kangari: A descriptive study of the Kangra Valley dialect of Himachal Pradesh.* Hoshiarpur: Vishveshvaranand Vishva Bandhu Institute of Sanskrit and Indological Studies, Panjab University.

Sharma, S. R. (1979). Phonological structure of Spiti. *Linguistics of the Tibeto-Burman Area, 4* (2), 83–110.

Shaw, P. A. (1991). Consonant harmony systems: The special status of coronal harmony. In C. Paradis, & J.-F. Prunet (Eds.), *The special status of coronals: Internal and external evidence* (pp. 125–157). San Diego: Academic Press.

Shepherd, G., & Shepherd, B. (1971). *Magar phonemic summary.* Tibeto-Burman Phonemic Summaries, 8. Kirtipur: Summer Institute of Linguistics and Tribhuvan University.

Shukla, S. (1981). *Bhojpuri grammar.* Washington, DC: Georgetown University Press.

Sibanda, G. (2004). *Verbal phonology and morphology of Ndebele.* Doctoral dissertation. University of California, Berkeley.

Simonsen, H. G., Moen, I., & Cowen, S. (2008). Norwegian retroflex stops in a cross linguistic perspective. *Journal of Phonetics, 36* (2), 385–405.

Singh, I. (1975). *Manipuri phonetic reader.* Mysore: Central Institute of Indian Languages.

Singh, J. D. (1970). *A descriptive grammar of Bangru.* Kurukshetra, Haryana: Kurukshetra University.

Sinha, K. P. (1974). Bishnupriya Manipuri: A descriptive sketch. *Indian Linguistics, 35*, 185–189.

Sinha, K. P. (1981). *The Bishnupriya Manipuri language.* Calcutta: Firma KLM Private Limited.

Soundararaj, J., & Soundararaj, O. (1999). *A phonological summary of Bison Horn Madiya.* Unpublished manuscript. Indian Evangelical Mission.

Sprigg, R. K. (2002). *Balti-English English-Balti dictionary.* London & New York: RoutledgeCurzon.

Sreedhar, M. V. (1976). *Sema phonetic reader.* Mysore: Central Institute of Indian Languages.

Sreedhar, M. V. (1979). Phonemes of Rongmei Naga. *Indian Linguistics, 40*, 41–48.

Sreekantaiya, T. N. (1954). Affricates in Kannada speech. *Indian Linguistics, 14*, 83–90.

Sridhar, S. N. (1990). *Kannada.* London & New York: Routledge.

Srivastava, G. P. (1968). Bahelia phonology. *Indian Linguistics, 29*, 67–79.

Stampe, D. L. (1966). Recent Work in Munda Linguistics IV. *International Journal of American Linguistics, 32* (4), 390-397.

Steever, S. B. (1998a). Gondi. In S. B. Steever, *The Dravidian Languages* (pp. 270–297). London & New York: Routledge.

Steever, S. B. (1998b). Introduction to the Dravidian languages. In S. B. Steever (Ed.), *The Dravidian languages* (pp. 1–39). London & New York: Routledge.

Steever, S. B. (1998c). Kannada. In S. B. Steever, *The Dravidian languages* (pp. 129–157). London & New York: Routledge.

Steever, S. B. (1998d). Malto. In S. B. Steever, *The Dravidian languages* (pp. 359–387). London & New York: Routledge.

Steingass, F. J. (1892). *A Comprehensive Persian-English dictionary, including the Arabic words and phrases to be met with in Persian literature.* London: Routledge & K. Paul. Online URL: http://dsal.uchicago.edu/dictionaries/steingass/.

Stenzler, A. F. (1997). *Primer of the Sanskrit language* (2nd revised edition ed.). (R. Söhnen-Thieme, Trans.) London: School of Oriental and African Studies, University of London.

Steriade, D. (1987). Redundant values. In A. Bosch, B. Need, & E. Schiller (Ed.), *Papers from the 23rd Annual Regional Meeting of the Chicago Linguistic Society* (pp. 339–362). Chicago: Chicago Linguistics Society.

Steriade, D. (2001). Directional asymmetries in place assimilation: A perceptual account. In E. Hume, & K. Johnson (Eds.), *The role of speech perception in phonology* (pp. 219-250). New York: Academic Press.

Strahm, E., & Maibaum, A. (1971). *Jirel phonemic summary.* [Kirtipur]: Summer Institute of Linguistics and Tribhuvan University.

Subbaiah, G. (1986). *Kota phonetic reader.* Mysore: Central Institute of Indian Languages.

Subrahmanyam, P. S. (1968). *A descriptive grammar of Gondi.* Annamalainagar: Annamalai University.

Subrahmanyam, P. S. (1983). *Dravidian comparative phonology.* Annamalainagar: Annamalai University.

Subrahmanyam, P. S. (1998). Kolami. In S. B. Steever, *The Dravidian languages* (pp. 301–327). London & New York: Routledge.

Suzuki, K. (1998). *A typological investigation of dissimilation.* Doctoral dissertation. University of Arizona.

Taylor, D. (1970). Tamang segmental synopsis. In A. Hale, & K. L. Pike (Eds.), *Tone systems of Tibeto-Burman languages of Nepal 1: Studies on tone and phonological segments* (pp. 237–57). Urbana: University of Illinois Press.

Thirumalai, M. S. (1972). *Thaadou phonetic reader.* Mysore: Central Institute of Indian Languages.

Thurgood, G. (2003). A subgrouping of the Sino-Tibetan languages: The interaction between language contact, change, and inheritance. In G. Thurgood, & R. J. LaPolla (Eds.), *The Sino-Tibetan languages* (pp. 3–21). London & New York: Routledge.

Tiffou, É., & Pesot, J. (1989). *Contes du Yasin: Intruduction au Bourouchaski du Yasin avec grammaire et dictionnaire analytique.* Paris: Peeters/Selaf.

Tikkanen, B. (1999). Archaeological-linguistic correlations in the formation of retroflex typologies and correlating areal features in South Asia. In R. Blench, & M. Spriggs (Eds.), *Archaeology and language IV: Language change and cultural transformation* (pp. 138-148). London & New York: Routledge.

Toba, S. (1984). *Khaling.* Tokyo: Institute for the Study of Languages and Cultures of Asian and Africa, Tokyo University of Foreign Studies.

Toba, S., & Toba, I. (1972). *Khaling phonemic summary.* Tibeto-Burman Phonemic Summaries, 12. Kirtipur: Summer Institute of Linguistics and Institute of Nepal Studies, Tribhuvan University.

Tolsma, G. J. (2006). *A grammar of Kulung.* Leiden & Boston: Brill.

Trail, R. L. (1970). *The grammar of Lamani.* Norman, Oklahoma: Summer Institute of Linguistics.

Trail, R. L., & Cooper, G. R. (1999). *Kalasha dictionary – with English and Urdu.* Islamabad: National Institute of Pakistan Studies, Quaid-i-Azam University and Summer Institute of Linguistics.

Trivedi, G. M. (1991). *Descriptive grammar of Byansi: A Bhotiya language.* Calcutta: Anthropological Survey of India, Government of India.

Turin, M. (2004). The phonology of Thangmi: A Tibeto-Burman language of Nepal. *Journal of Asian and African Studies, 67*, 63–103.

Turner, R. L. (1931). *A comparative and etymological dictionary of the Nepali language.* London: K. Paul, Trench, Trubner. Online URL: http://dsal1.uchicago.edu/dictionaries/turner/.

Turner, R. L. (1962–1966). *A comparative dictionary of Indo-Aryan languages.* London: Oxford University Press. Online URL: http://dsal.uchicago.edu/dictionaries/soas/.

Turner, R. L. (1969). *A comparative dictionary of Indo-Aryan languages. Indexes.* London: Oxford University Press.

Upadhyaya, U. P. (1972). *Kannada phonetic reader.* Mysore: Central Institute of Indian Languages.

van Driem, G. (1987). *A grammar of Limbu.* Berlin & New York: Mouton de Gruyter.

van Driem, G. (1992). *The grammar of Dzongkha.* Thimphu, Bhutan: Dzongkha Development Commission, Royal Government of Bhutan.

van Driem, G. (1993). *A grammar of Dumi.* Berlin & New York: Mouton de Gruyter.

van Riezen, I. (2000). *An outline of Kumauni phonology*. Unpublished manuscript.

Varkey, V. W., & Vinod, A. (2003). *Rathwi Bareli phonemic summary*. Unpublished manuscript. New Life Computer Institute.

Varma, G. S. (1970). *Vaagri Boli: An Indo-Aryan language.* Annamalainagar: Annamalai University.

Varma, G. S. (1978a). *Kurumba Kannada.* Annamalainagar: Annamalai University.

Varma, G. S. (1978b). *Yerukala dialect.* Annamalainagar: Annamalai University.

Verma, M. K. (2003). Bhojpuri. In G. Cardona, & D. Jain, *The Indo-Aryan languages* (pp. 515–537). London & New York: Routledge.

Verma, S. (2003). Magahi. In G. Cardona, & D. Jain, *The Indo-Aryan languages* (pp. 498–514). London & New York: Routledge.

Vesalainen, O., & Vesalainen, M. (1976). *Lhomi phonemic summary.* Kathmandu: Summer Institute of Linguistics and Institute of Nepal and Asian Studies.

Wali, K. (2005). *Marathi.* Muenchen: Lincom Europa.

Wali, K., & Koul, O. N. (1997). *Kashmiri: A cognitive-descriptive grammar.* London & New York: Routledge.

Walker, R. (1998). *Nasalization, neutral segments, and opacity effects.* Doctoral dissertation. University of California, Santa Cruz. [Published, New York: Garland, 2000.].

Walker, R. (2000a). Long-distance consonantal identity effects. In Billerey, & Lillehaugen (Ed.), *WCCFL 19 Proceedings* (pp. 532–545). Somerville, MA: Cascadilla Press.

Walker, R. (2000b). Yaka nasal harmony: spreading or segmental correspondence? In L. J. Conathan, J. Good, D. Kavitskaya, A. B. Wulf, & A. C. Yu (Ed.), *Proceedings of the*

26th Annual Meeting of the Berkeley Linguistics Society (pp. 321 – 332). Berkeley: University of California, Berkeley Linguistics Society.

Walker, R. (2001). Consonantal correspondence. In R. Kirchner, J. Pater, & W. Wikeley (Ed.), *Proceedings of the Workshop on the Lexicon in Phonetics and Phonology* (pp. 73–84). Edmonton: University of Alberta, Department of Linguistics.

Walker, R., & Mpiranya, F. (2005). On triggers and opacity in coronal harmony. (R. T. Cover, & Y. Kim, Eds.) *Proceedings of Berkeley Linguistics Society, 31*, 383-394.

Walker, R., Byrd, D., & Mpiranya, F. (2008). An articulatory view of Kinyarwanda coronal harmony. *Phonology, 25*, 499-535.

Watkins, J. W. (2001). Burmese. *Journal of the International Phonetic Association, 31* (2), 291–295.

Watters, D. E. (2002). *A grammar of Kham.* Cambridge: Cambridge University Press.

Watters, D. E. (2003). Kham. In G. Thurgood, & R. J. LaPolla (Eds.), *The Sino-Tibetan languages* (pp. 683–704). London & New York: Routledge.

Watters, D. E. (2006). Notes on Kusunda grammar: A language isolate of Nepal. *Himalayan Linguistics Archive, 3*, 1–182.

Watters, S. A. (1996). *A preliminary study of prosody in Dzongkha*. MA thesis. University of Texas at Arlington.

Werker, J. F., Gilbert, J. H., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development, 52* (1), 349-355.

Wheatley, J. K. (1992). Burmese. In W. Bright (Ed.), *International Encyclopedia of Linguistics* (pp. 206–210). New York & Oxford: Oxford University Press.

Wheatley, J. K. (2003). Burmese. In G. Thurgood, & R. J. LaPolla (Eds.), *The Sino-Tibetan languages* (pp. 195–207). London & New York: Routledge.

Whitney, W. D. (1993 [1889]). *Sanskrit grammar including both the classical language and the older dialects of Veda and Brahmana* (2nd ed.). Cambridge, MA: Harvard University Press.

Wilde, C. P. (2001). *Preliminary phonological analysis of the Limi dialect of Humla Bhotia.* MA thesis. University of Helsinki.

Wilde, C. P. (2008). *A sketch of the phonology and grammar of Rājbanshi.* Doctoral dissertation. Helsinki: University of Helsinki.

Wilson, C. (2000). *Targeted constraints: an approach to contextual neutralization in optimality theory.* Doctoral dissertation. John Hopkins University.

Wilson, C. (2001). Consonant cluster neutralisation and targeted constraints. *Phonology, 18* (1), 147-197.

Winfield, W. W. (1928). *A grammar of the Kui language.* Calcutta: Asiatic Society of Bengal.

Witzel, M. (1999). Substrate languages in Old Indo-Aryan (R̥gvedic, Middle and Late Vedic). *Electronic Journal of Vedic Studies, 5* (1), 1–67.

Wright, R. (2004). A review of perceptual cues and cue robustness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology* (pp. 34-57). Cambridge: Cambridge University Press.

Yadav, R. (1996). *A reference grammar of Maithili.* Berlin & New York: Mouton de Gruyter.

Yadav, R. (2003). Maithili. In G. Cardona, & D. Jain, *The Indo-Aryan languages* (pp. 477–497). London & New York: Routledge.

Zide, N. H. (1960). *Korku phonology and morphophonemics*. Doctoral dissertation. University of Pennsylvania.

Zide, N. H. (2008). Korku. In G. D. Anderson (Ed.), *The Munda languages* (pp. 256–298). London & New York: Routledge.

Zoller, C. P. (2005). *A grammar and dictionary of Indus Kohistani, volume 1: Dictionary.* Berlin & New York: Mouton de Gruyter.

Zvelebil, K. (1970). *Comparative Dravidian phonology.* The Hague: Mouton.

Zvelebil, K. (1973). *The Irula language.* Wiesbaden: Otto Harrassowitz.

Appendix A
List of South Asian languages surveyed for retroflexion

This appendix provides a complete list of all languages and language varieties included in the survey of retroflexion in South Asia (Chapter 2). Each language is listed alphabetically by name along with its ISO 639-3 code and genetic classification, as found in the sixteenth edition of the *Ethnologue* (Lewis, 2009). The descriptive sources consulted for each language are also listed. The following abbreviations are used: AA = Austro-Asiatic; Dr = Dravidian; IA = Indo-Aryan; IIr = Indo-Iranian; TB = Tibeto-Burman.

Language Name	ISO	Classification	Descriptive Data Sources
Aiton	aio	Tai-Kadai	(Morey, 2005)
Angami, Khonoma	njm	TB, Kuki-Chin-Naga	(Blankenship, Ladefoged, Bhaskararao, & Chase, 1993; Ravindran, 1974)
Ao, Chungli	njo	TB, Kuki-Chin-Naga	(Gowda, 1972)
Ao, Mongsen	njo	TB, Kuki-Chin-Naga	(Coupe, 2007)
Apatani	apt	TB, Tani	(Abraham, 1985)
Asamiya (Assamese)	asm	IIr, IA, Eastern	(Goswami G. C., 1966; Goswami & Tamuli, 2003)
Athpariya (Athpare)	aph	TB, Kiranti	(Ebert, 1997a)
Awadhi	awa	IIr, IA, East-Central	(Saksena, 1937)
Badaga	bfq	Dr, Southern	(Hockings & Pilot-Raichoor, 1992)
Bagri	bgq	IIr, IA, Central	(Gusain, 2000)
Balochi	bal	IIr, Iranian, Western	(Elfenbein, 1997a)
Balti	bft	TB, Tibetic, Tibetan	(Lobsang, 1995; Sprigg, 2002; Rangan, 1975)
Bangla (Bengali)	ben	IIr, IA, Eastern	(Bhattacharya K. , 1988; Chatterji, 1970 [1926]; Dasgupta, 2003)
Bareli, Pawri	bfb	IIr, IA, Central	(Immanuel & Jane, 2003)
Bareli, Rathwi	bgd	IIr, IA, Central	(Varkey & Vinod, 2003)
Belhariya (Belhare)	byw	TB, Kiranti	(Bickel, 2003)
Bhatri	bgw	IIr, IA, Eastern	(Kirivasan & Amirthamary, 2000)
Bhil, Dungra	duh	IIr, IA, Central	(Mathew & Susan, 2000)

Language Name	ISO	Classification	Descriptive Data Sources
Bhojpuri	bho	IIr, IA, Eastern	(Shukla, 1981; Verma M. K., 2003)
Bhumij (dialect of Mundari)	unr	AA, Munda, Northern	(Ramaswami, 1992)
Bishnupriya	bpy	IIr, IA, Eastern	(Sinha, 1974; 1981)
Bodo (Boro)	brx	TB, Jingpho-Konyak-Bodo	(Bhattacharya P. C., 1977)
Brahui	brh	Dr, Northern	(Elfenbein, 1997b; 1998)
Brijia (dialect of Asuri)	asr	AA, Munda, Northern	(Sahu, 1980)
Brokskat	bkk	IIr, IA, Northwestern, Dardic	(Ramaswami, 1975; Bashir, 2003)
Bundeli	bns	IIr, IA, Central	(Jaiswal, 1962)
Burmese	mya	TB, Lolo-Burmese	(Watkins, 2001; Wheatley, 1992; 2003)
Burushaski	bsk	Isolate	(Anderson, 1997)
Byangsi	bee	TB, Western Himalayish	(Trivedi, 1991)
Camling	rab	TB, Kiranti	(Ebert, 1997b; 2003)
Chantyal	chx	TB, Tamangic	(Noonan, 2003a)
Chepang	cdm	TB, Kham-Magar-Chepang-Sunwari	(Caughley, 1969; 1970)
Chokri	nri	TB, Kuki-Chin-Naga	(Bielenberg & Nienu, 2001)
Dameli	dml	IIr, IA, Northwestern, Dardic	(Bashir, 2003)
Darai	dry	IIr, IA, Unclassified	(Kotapish & Kotapish, 1973)
Deori	der	TB, Jingpho-Konyak-Bodo	(Jacquesson, 2005; Goswami U. , 1994)
Desiya (Desiya Oryia)	dso	IIr, IA, Eastern	(Mathews, 2003)
Dhanki	dhn	IIr, IA, Central	(Kulkarni, 1976)
Dhanwar	dhw	IIr, IA, Unclassified	(Kuegler & Kuegler, 1974)
Dhimal	dhi	TB, Tibetic	(Cooper, 1999; King, 2008)
Dimasa	dis	TB, Jingpho-Konyak-Bodo	(Misra, 1986)
Dhivehi (Maldivian)	div	IIr, IA, Sinhalese-Maldivian	(Cain & Gair, 2000)
Dogri	dgo	IIr, IA, Northern	(Ghai, 1991)
Domaaki	dmk	IIr, IA, Northwestern, Dardic	(Lorimer, 1939)

Language Name	ISO	Classification	Descriptive Data Sources
Dumi	dus	TB, Kiranti	(van Driem, 1993)
Dzongkha	dzo	TB, Tibetic, Tibetan	(Mazaudon & Michailovsky, 1988; van Driem, 1992; Watters S. A., 1996)
Gadaba, Mudhili (Konekor)	gau	Dr, Central	(Bhaskararao, 1980; 1998; Rao & Patnaik, 1992)
Gadaba, Ollari	gdb	Dr, Central	(Bhattacharya S. , 1957)
Gallong (Galo Adi)	adl	TB, Tani	(Gupta, 1963)
Garasia (Rajput Garasia)	gra	IIr, IA, Central	(Patel, 1999)
Garhwali	gbm	IIr, IA, Northern	(Chandrasekhar, 1969)
Garo (Mande)	grt	TB, Jingpho-Konyak-Bodo	(Burling, 2003)
Gondi, Muria	emu, mut	Dr, South-Central	(Steever, 1998a)
Gondi, Southern (Adilabad)	ggo	Dr, South-Central	(Subrahmanyam P. S., 1968; Lincoln, 1969; Kurian & Kurian, 2000)
Great Andamanese	apq	Andamanese	(Manoharan, 1989; Abbi, 2006)
Gtaʔ	gaq	AA, Munda, Southern	(Anderson, 2008a)
Gujarati	guj	IIr, IA, Central	(Mistry, 1997; Cardona & Suthar, 2003)
Gurung	gvr, ggn	TB, Tibetic, Tamangic	(Glover, 1969; 1970)
Gutob (Gadaba)	gbj	AA, Munda, Southern	(Griffiths, 2008; Rajan & Rajan, 2001)
Hakha Lai (Haka Chin)	cnh	TB, Kuki-Chin-Naga	(Peterson D. A., 2003)
Haryanvi (Bangru)	bgc	IIr, IA, Central	(Singh J. D., 1970)
Hayu (Wayu)	vay	TB, Kham-Magar-Chepang-Sunwari	(Michailovsky, 2003)
Hindi	hin	IIr, IA, Central	(Kaye, 1997; Ohala, 1983; 1994; Shapiro, 2003)
Hmar	hmr	TB, Kuki-Chin-Naga	(Baruah & Bapui, 1996)
Ho	hoc	AA, Munda, Northern	(Anderson, Osada, & Harrison, 2008)
Humla, Limi dialect	hut	TB, Tibetic, Tibetan	(Wilde, 2001)
Irula	iru	Dr, Southern	(Periyalwar, 1979; Zvelebil K. V., 1973)
Jarawa	anq	Andamanese	(Abbi, 2006)
Jaunsari	jns	IIr, IA, Northern	(Satish, 1990)
Jero (Jerung)	jee	TB, Kiranti	(Opgenort, 2005)
Jinghpo	kac	TB, Jingpho-Konyak-Bodo	(Qingxia & Diehl, 2003)
Jirel	jul	TB, Tibetan	(Strahm & Maibaum, 1971)

Language Name	ISO	Classification	Descriptive Data Sources
Juang	jun	AA, Munda, Southern	(Mathew & Mathew, 2003; Matson, 1964; Patnaik, 2008)
Kagate	syw	TB, Tibetan	(Höhlig & Hari, 1976)
Kalami (Kalam Kohistani)	gwc	IIr, IA, Northwestern, Dardic	(Baart, 1997; 2004; Bashir, 2003)
Kalasha	kls	IIr, IA, Northwestern, Dardic	(Heegård & Mørch, 2004; Mørch & Heegård, 1997; Trail & Cooper, 1999)
Kangri	xnr	IIr, IA, Northern	(Sharma S. , 1974)
Kannada	kan	Dr, Southern	(Sridhar, 1990; Steever, 1998c; Upadhyaya, 1972)
Karbi (Mikir)	mjw	TB, Mikir	(Jeyapaul, 1987)
Kasaba (dialect of Irula)	iru	Dr, Southern	(Pillai, 1976)
Kashmiri	kas	IIr, IA, Northwestern, Dardic	(Bhat R. , 1987; Koul, 2003; Wali & Koul, 1997; Handoo, 1973)
Khaling	klr	TB, Kiranti	(Toba & Toba, 1972; Toba, 1984)
Kham	kjl	TB, Kham-Magar-Chepang-Sunwari	(Watters D. E., 2002; 2003)
Khamyang	ksu	Tai-Kadai	(Morey, 2005)
Kharia	khr	AA, Munda, Southern	(Biligiri, 1965; Peterson J. , 2008)
Khasi	kha	AA, Mon-Khmer	(Nagaraja, 1985; 1989)
Khowar	khw	IIr, IA, Northwestern, Dardic	(Bashir, 2003)
Kinnauri	kfk	TB, Western Himalayish	(Sharma D. D., 1988)
Kodaku	ksz	AA, Munda, Northern	(Kuriakkose & Liju, 2008)
Kodava (Kodagu, Coorg)	kfa	Dr, Southern	(Balakrishnan, 1976; 1977; Ebert, 1996)
Kohistani, Indus	mvy	IIr, IA, Northwestern, Dardic	(Hallberg & Hallberg, 1999; Bashir, 2003; Zoller, 2005)
Koi (Koyi, Kohi)	kkt	TB, Kiranti	(Lahaussois, 2009)
Kok Borok (Tripuri)	trp	TB, Jingpho-Konyak-Bodo	(Karapurkar, 1972; 1976)
Kolami, Northwestern (Wardha dialect)	kfb	Dr, Central	(Emeneau, 1961; Subrahmanyam P. S., 1998)
Konda (Kubi)	kfc	Dr, South-Central	(Krishnamurti, 1969; Krishnamurti & Benham, 1998)
Konkani	knn	IIr, IA, Southern	(Katre, 1966; Almeida, 1985; Miranda, 2003)
Koraga, Korra	kfd	Dr, Southern	(Bhat D. N., 1971)

Language Name	ISO	Classification	Descriptive Data Sources
Koraga, Mudu	vmd	Dr, Southern	(Bhat D. N., 1971)
Korku	kfq	AA, Munda, Northern	(Zide, 1960; 2008)
Korwa	kfp	AA, Munda, Northern	(George & Joseph, 2008)
Kota	kfe	Dr, Southern	(Subbaiah, 1986)
Kui	kxu	Dr, South-Central	(Winfield, 1928)
Kulung	kle	TB, Kiranti	(Tolsma, 2006)
Kumauni	kfy	IIr, IA, Northern	(Apte & Pattanayak, 1967; van Riezen, 2000)
Kundal Shahi	??	IIr, IA, Northwestern?	(Rehman & Baart, 2005)
Kurtöp (Kurtokha)	xkz	TB, Tibetan	(Hyslop, 2008)
Kurumba Kannada	kfi	Dr, Southern	(Varma, 1978a; Ernest & Ernest, 2000)
Kurumba, Betta	xub	Dr, Southern	(Selvaraj & Selvaraj, 2003)
Kurux	kru	Dr, Northern	(Hahn, 1911; Pfeiffer, 1972)
Kusunda	kgg	Isolate	(Watters D. E., 2006)
Kuvi	kxv	Dr, South-Central	(Reddy, Upadhyaya, & Reddy, 1974; Israel, 1979)
Kyerung	kgy	TB, Tibetan	(Huber, 2005)
Ladakhi	lbj	TB, Tibetan	(Koshal, 1976; 1979)
Lamani (Lambadi, Banjara)	lmn	IIr, IA, Central	(Trail, 1970)
Lepcha	lep	TB, Tibeto-Kanauri	(Plaisier, 2003; 2007)
Lhomi	lhm	TB, Tibetan	(Vesalainen & Vesalainen, 1976)
Limbu	lif	TB, Kiranti	(van Driem, 1987)
Lisu	lis	TB, Lolo-Burmese	(Bradley, 2003)
Lotha	njh	TB, Kuki-Chin-Naga	(Acharya K. P., 1975; 1983)
Lushai (Mizo)	lus	TB, Kuki-Chin-Naga	(Burling, 1957)
Magahi	mag	IIr, IA, Eastern	(Verma S. , 2003)
Magar, Eastern	mgp	TB, Kham-Magar-Chepang-Sunwari	(Shepherd & Shepherd, 1971)
Maithili	mai	IIr, IA, Eastern	(Yadav, 1996; 2003; Jha, 2001)
Malayalam	mal	Dr, Southern	(Asher & Kumari, 1997; Kumari, 1972)
Malto	kmj, mjt	Dr, Northern	(Mahapatra, 1979; Steever, 1998d; Das A. S., 1973)
Manangba (Manange)	nmm	TB, Tamangic	(Hildebrandt, 2004)
Mao Naga	nbi	TB, Kuki-Chin-Naga	(Giridhar, 1994)
Marathi	mar	IIr, IA, Southern	(Pandharipande, 1997; 2003; Wali, 2005)

Language Name	ISO	Classification	Descriptive Data Sources
Maria, Hill (Abujhmaria)	mrr	Dr, South-Central	(Natarajan, 1985)
Maria, Dandami (Bison Horn Maria)	daq	Dr, South-Central	(Soundararaj & Soundararaj, 1999)
Marwari	rwr	IIr, IA, Central	(Gusain, 2004)
Meithei (Manipuri)	mni	TB, Meitei	(Bhat & Ningomba, 1997; Chelliah, 1997; 2003; Singh I. , 1975)
Mewati	wtm	IIr, IA, Central	(Gusain, 2003)
Mishmi, Digaro dialect	mhu	TB, Tani	(Sastry G. D., 1984a; 1984b)
Mising (Miri)	mrg	TB, Tani	(Prasad, 1991)
Mundari	unr	AA, Munda, Northern	(Cook, 1965; Gumperz, 1957; Kobayashi & Murmu, 2008; Osada, 2008)
Nar Phu	npa	TB, Tamangic	(Noonan, 2003b)
Nepali	nep	IIr, IA, Northern	(Acharya J. , 1991; Riccardi, 2003; Khatiwada, 2009)
Newar, Kathmandu (Nepāl Bhāśā)	new	TB, Mahakiranti, Newari	(Hale, 1970; Hargreaves, 2003)
Newar, Dolakha	new	TB, Mahakiranti, Newari	(Genetti, 2003; 2007)
Nicobarese, Car	caq	AA, Mon-Khmer	(Das A. R., 1977)
Nicobarese, Central	ncb	AA, Mon-Khmer	(Radhakrishnan, 1981)
Onge	oon	Andamanese	(Abbi, 2006)
Oriya	ori	IIr, IA, Eastern	(Ray, 2003)
Palula	phl	IIr, IA, Northwestern, Dardic	(Liljegren, 2008; Bashir, 2003)
Paniya	pcg	Dr, Southern	(Daniel & Stephan, 2003)
Panjabi, Eastern	pan	IIr, IA, Central	(Bhatia, 1993; Jain, 1934; Malik, 1995; Shackle, 2003)
Pardhi (Bahelia)	pcl	IIr, IA, Central	(Srivastava, 1968)
Parenga (Gorum)	pcj	AA, Munda, Southern	(Anderson & Rau, 2008)
Parji (Duruwa)	pci	Dr, Central	(Burrow & Bhattacharya, 1953)
Pashto	pus	IIr, Iranian, Eastern	(Elfenbein, 1997c)
Pengo	peg	Dr, South-Central	(Burrow & Bhattacharya, 1970)
Phake	phk	Tai-Kadai	(Morey, 2005)
Phom	nph	TB, Jingpho-Konyak-Bodo	(Burling & Phom, 1998)

Language Name	ISO	Classification	Descriptive Data Sources
Rabha	rah	TB, Jingpho-Konyak-Bodo	(Joseph, 2007)
Rājbanshi	rjb	IIr, IA, Eastern	(Wilde, 2008)
Remo (Bonda, Bondo)	bfw	AA, Munda, Southern	(Alexander & Hannah, 2000; Anderson & Harrison, 2008a; Fernandez, 1968)
Rongmei Naga	nbu	TB, Kuki-Chin-Naga	(Sreedhar, 1979)
Sadri (Sadani)	sck	IIr, IA, Eastern	(Jordan-Horstmann, 1969)
Sanskrit	san	IIr, IA, Old Indo-Aryan	(Whitney, 1993 [1889]; Cardona, 2003)
Santali	sat	AA, Munda, Northern	(Ghosh, 2008; Neukom, 2001; Sebeok, 1943)
Saurashtra	saz	IIr, IA, Central	(Norihiko, 1991)
Sawi	sdg	IIr, IA, Northwestern, Dardic	(Bashir, 2003)
Seke (Tangbe dialect)	skj	TB, Tamangic	(Honda, 2003)
Sema (Sumi Naga)	nsm	TB, Kuki-Chin-Naga	(Sreedhar, 1976)
Seraiki	skr	IIr, IA, Northwestern	(Shackle, 1976; 2003)
Shekawati	swv	IIr, IA, Central	(Gusain, 2001)
Sherpa	xsr	TB, Tibetan	(Gordon & Schoettelndreyer, 1970; Kelly, 2004)
Shina, Gilgit	scl	IIr, IA, Northwestern, Dardic	(Radloff, 1999; Bashir, 2003)
Shina, Kohistani	plk	IIr, IA, Northwestern, Dardic	(Schmidt & Kohistani, 2008)
Sindhi	snd	IIr, IA, Northwestern	(Khubchandani, 2003; Nihalani, 1999)
Sinhala (Sinhalese)	sin	IIr, IA, Sinhalese-Maldivian	(Gair, 2003)
Sora	srb	AA, Munda, Southern	(Anderson & Harrison, 2008b)
Spiti	spt	TB, Tibetan	(Sharma S. R., 1979)
Sunwar	suz	TB, Kham-Magar-Chepang-Sunwari	(Borchers, 2008; Bieri & Schulze, 1969; 1970)
Tamang, Eastern	taj	TB, Tamangic	(Mazaudon, 2003; Poudel, 2006; Taylor, 1970)
Tamil, Modern	tam	Dr, Southern	(Annamalai & Steever, 1998; Christdas, 1988; Rajaram, 1972; Keane, 2004)
Tamil, Old Literary	tam	Dr, Southern	(Lehmann, 1998)
Tangkhul	nmf	TB, Kuki-Chin-Naga	(Arokianathan, 1987)
Telugu	tel	Dr, South-Central	(Krishnamurti, 1998; Sastry J. V., 1972)
Thado (Thadou)	tcz	TB, Kuki-Chin-Naga	(Krishan, 1980; Thirumalai, 1972)

Language Name	ISO	Classification	Descriptive Data Sources
Thakali	ths	TB, Tamangic	(Hari, 1969; 1970)
Thangmi	thf	TB, Western Himalayish	(Turin, 2004)
Tharu	the, thl, thq, thr	IIr, IA, Eastern	(Boehm, 1998; Gnanasekaran & Sheeba, 2003)
Thulung	tdh	TB, Kiranti	(Allen N. J., 1975; Lahaussois, 2002; 2003)
Tibetan, Lhasa	bod	TB, Tibetan	(DeLancey, 2003b; Denwood, 1999)
Toda	tcx	Dr, Southern	(Sakthivel, 1976; 1977; Shalev, Ladefoged, & Bhaskararao, 1993)
Torwali	trw	IIr, IA, Northwestern, Dardic	(Lunsford, 2001; Bashir, 2003)
Tshangla	tsj	TB, Tibetic	(Andvik, 2003)
Tulu	tcy	Dr, Southern	(Bhat D. N., 1967; 1998; Bhatt, 1971)
Urali	url	Dr, Southern	(Lal, 1991)
Urdu	urd	IIr, IA, Central	(Schmidt, 2003)
Vaagri Boli	vaa	IIr, IA, Unclassified	(Varma, 1970)
Wambule	wme	TB, Kiranti	(Opgenort, 2004)
Yerava (Ravula)	yea	Dr, Southern	(Mallikarjun, 1993)
Yerukala	yeu	Dr, Southern	(Varma, 1978b)

Appendix B
Coronal co-occurrence statistics

The survey of retroflex consonant harmony in South Asia, reported in Chapter 3, makes use of statistical information on coronal co-occurrence patterns in South Asian languages. In most cases, the information is abbreviated by collapsing phonological categories (e.g., ignoring laryngeal distinctions), omitting expected frequencies, or omitting numerical values altogether in favour of schematic representations. This appendix provides details, omitted in Chapter 3, concerning coronal co-occurrence statistics in the following 41 languages.

For each language, the following information is provided:

Language Name: Languages are presented alphabetically by name. Where appropriate, popular alternative names are listed in parentheses.

Classification: The genetic classification of each language is identified. This includes the language family and any major subgroup(s) to which it belongs.

Descriptive Sources: Descriptive sources refer to phonological descriptions that were consulted in establishing the phonemic inventory of each language.

Consonant Phonemes: The consonant phonemes of each language are listed in order to clarify what segments were and were not included in the statistical analysis. Marginal phonemes are listed in parentheses. Such phonemes may be limited to loanwords or to a particular dialect, or their phonemic status may not be recognized in all of the data sources.

Lexical Data Source: This refers to the source of lexical data over which counts were made. It may be a published dictionary, a vocabulary list in a published grammar, an online dictionary or database, or an unpublished database. In most cases, this source is distinct from the descriptive source(s) consulted for phonological information.

Search Domain: This refers to the domain within which coronal co-occurrence counts were made. In most cases, the domain was limited to headwords containing word-initial C_1VNC_2 sequences. In a few cases, counts were made over shorter or longer domains. In some cases, separate counts were made of headwords and unique roots. In those cases, the domains are distinguished as "Search Domain 1" and "Search Domain 2".

Segment Class: This refers to the class of segments that were included in the counts. At the very least, this class included coronal plosives. In most cases it also included any retroflex sonorants that were distinguished in the lexical data source(s). Where appropriate, coronal affricates and fricatives were also included. Where more than one class was examined, the classes are identified as “Segment Class 1” and “Segment Class 2”.

Observed Counts: For each unique search domain or segment class, the observed counts for each individual $C_1 \ldots C_2$ pair are listed in a table. In each table, the vertical axis represents C_1 and the horizontal axis represents C_2.

Expected Frequencies and O/E Ratios: For each unique count, expected (E) frequencies and O/E ratios are reported in a separate table. These tables typically collapse some phonemic categories by ignoring laryngeal distinctions or some manner distinctions. In most cases, these tables correspond to tables presented in Chapter 3, where one or more details have been omitted in order to streamline the presentation.

Bangla (Bengali)

Classification:	Indo-Aryan, Eastern					
Descriptive Source(s):	Bhattacharya (1988); Chatterji (1970 [1926]); Dasgupta (2003)					
Consonant Phonemes:	p	t	ʈ	ʧ	k	
	p^h	t^h	ʈ^h	ʧ^h	k^h	
	b	d	ɖ	ʤ	g	
	b^h	d^h	ɖ^h	ʤ^h	g^h	
	(f)	(s)		ʃ		h
	(v)	(z)				
	m	n			ŋ	
		r	ɽ			
		l				
	(w)			(j)		
Lexical Data Source:	Biswas (2000)					
Search Domain:	headwords containing $\#C_1V(N)C_2$ sequences					
Segment Class:	$\{C_1, C_2\}$ = coronal plosives and retroflex sonorants					

Table B-1 Bangla: Observed values for each $C_1 \ldots C_2$ pair (n = 357)

$C_1 \backslash C_2$	t	t^h	d	d^h	ʈ	ʈ^h	ɖ	ɖ^h	ɽ
t	52	8	46	-	5	-	4	-	50
t^h	8	1	-	-	-	-	-	-	6
d	27	-	20	4	2	-	10	-	24
d^h	13	-	3	3	1	-	-	-	13
ʈ	-	-	-	-	23	-	-	-	2
ʈ^h	-	-	-	-	9	1	2	-	1
ɖ	-	-	-	-	6	-	4	-	1
ɖ^h	-	-	-	-	-	-	-	2	6
ɽ	-	-	-	-	-	-	-	-	-

Table B-2 Bangla: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner (n = 357)

$C_1 \backslash C_2$		t, t^h, d, d^h	ʈ, ʈ^h, ɖ, ɖ^h	ɽ
t, t^h, d, d^h	O	185	22	93
	E	155.5	58.0	86.6
	O/E	1.19	0.38	1.07
ʈ, ʈ^h, ɖ, ɖ^h	O	0	47	10
	E	29.5	11.0	16.4
	O/E	0.00	4.27	0.61

Burushaski

Classification: Isolate

Descriptive Source(s): Anderson (1997)

Consonant Phonemes:

p	t	ʈ		k	q	
pʰ	tʰ	ʈʰ		kʰ	qʰ	
b	d	ɖ		g	ɢ	
	ʦ	ʈʂ	ʧ			
	ʦʰ	ʈʂʰ	ʧʰ			
		ɖʐ	ʤ			
(f)	s	ʂ	ʃ	(x)		h
	z					
m	n			ŋ		
	r, l					
w		ɻ	j			

Lexical Data Source: Berger (1998b)

Search Domain: headwords containing $\#C_1VC_2$ sequences

Segment Class: $\{C_1, C_2\}$ = coronal obstruents (plosives, affricates, fricatives)

Table B-3 Burushaski: Observed values for each $C_1 \ldots C_2$ pair (n = 661)

$C_1 \backslash C_2$	t	tʰ	d	ʈ	ʈʰ	ɖ	ʦ	ʦʰ	ʧ	ʧʰ	ʤ	ʈʂ	ʈʂʰ	ɖʐ	s	z	ʃ	ʂ
t	52	-	1	2	-	8	2	-	2	-	3	1	-	-	21	3	5	11
tʰ	1	16	-	4	-	7	1	1	-	-	-	-	-	-	2	-	6	2
d	4	-	14	1	-	3	3	1	3	-	1	-	-	-	12	-	5	11
ʈ	-	-	-	25	-	1	1	-	-	-	-	-	-	-	2	-	-	-
ʈʰ	-	-	-	3	7	-	4	-	-	-	-	-	-	-	1	-	-	-
ɖ	-	-	-	-	-	36	-	-	-	-	-	-	-	-	6	-	-	-
ʦ	-	-	-	2	-	-	11	-	-	-	-	-	-	-	-	-	-	-
ʦʰ	-	-	-	3	-	-	2	13	-	-	-	1	-	-	-	-	-	-
ʧ	8	-	5	11	-	2	-	-	18	-	1	-	-	-	10	1	1	4
ʧʰ	5	-	-	6	-	1	-	-	-	5	-	-	-	-	1	-	2	6
ʤ	1	-	1	18	2	8	-	-	4	-	25	-	-	-	1	-	-	7
ʈʂ	1	-	-	6	-	-	-	-	-	-	-	4	-	-	-	-	-	-
ʈʂʰ	1	-	-	1	-	-	-	-	1	-	-	-	1	-	-	-	-	-
ɖʐ	1	-	-	1	-	2	-	-	-	-	-	-	-	5	-	-	-	1
s	10	2	7	4	-	1	1	-	2	-	4	10	-	-	11	1	-	-
z	1	-	2	2	-	-	-	-	-	-	-	-	-	-	-	12	-	-
ʃ	18	-	4	7	1	-	-	1	1	-	3	-	-	-	4	-	12	1
ʂ	6	-	3	2	-	3	-	-	-	-	-	-	-	-	-	-	-	19

Table B-4 Burushaski: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner; linear order of segments collapsed (n = 661)

C_1\\C_2		t, t^h, d	ʈ, ʈ^h, ɖ	ʦ, ʦ^h	ʧ, ʧ^h, ʤ	ʈʂ, ʈʂ^h, ɖʐ	s, z	ʃ	ʂ
t, t^h, d	O	88	25	8	29	4	60	38	33
	E	51.9	78.3	20.9	61.3	13.2	45.2	22.7	27.8
	O/E	1.70	0.32	0.38	0.47	0.30	1.33	1.67	1.19
ʈ, ʈ^h, ɖ	O		72	10	48	10	16	8	5
	E		23.4	14.0	51.4	9.7	30.5	18.2	17.1
	O/E		3.07	0.71	0.93	1.03	0.52	0.44	0.29
ʦ, ʦ^h	O			26	0	1	1	1	0
	E			2.0	13.1	2.6	8.6	4.7	5.0
	O/E			13.10	0.00	0.38	0.12	0.21	0.00
ʧ, ʧ^h, ʤ	O				53	1	19	7	17
	E				17.0	7.9	28.2	13.0	18.1
	O/E				3.12	0.13	0.67	0.54	0.94
ʈʂ, ʈʂ^h, ɖʐ	O					10	10	0	1
	E					0.8	5.7	2.9	3.4
	O/E					12.02	1.77	0.00	0.29
s, z	O						24	4	0
	E						9.3	10.2	11.0
	O/E						2.58	0.39	0.00
ʃ	O							12	1
	E							2.4	6.4
	O/E							4.92	0.16
ʂ	O								19
	E								3.1
	O/E								6.14

Note: While co-occurring obstruents that agree in manner of articulation are over-attested in Burushaski, the vast majority of them appear to be the product of reduplication (e.g., /tatal/ from /tal/ ‘flow slowly’; /ʈaʈaŋ/ from /ʈaŋ/ ‘cloudy’, /ʦáʦar/ from /ʦar/ ‘tear up, split’; /ʧaʧáq/ from /ʧaq/ ‘chew’; /ʈʂaʈʂaʈ/ from /ʈʂaʈ/ ‘closed gap’; /zazál/ from /zal/ ‘shake’; /ʃaʃál/ from /ʃal/ ‘shallow’; /ʂúʂu/ from /ʂu/ ‘scare away, drive off’). This fact, combined with relatively high observed counts for some disharmonic sequences, such as T-Ṭ and Č-Ṣ, raises doubts about the extent to which retroflex consonant harmony has contributed to the surface pattern, or if it has at all.

Dhivehi (Maldivian)

Classification: Indo-Aryan, Sinhalese-Maldivian

Descriptive Source(s): Cain & Gair (2000)

Consonant Phonemes:

p	t	ʈ	ʧ	k	
b	d	ɖ	ʤ	g	
^{m}b	^{n}d	^{n}ɖ		^{n}g	
f	s	ʂ			h
v	z				
m	n		(ɲ)		
	r				
	l	ɭ			
			j		

Lexical Data Source: Reynolds (2003)

Search Domain: headwords containing $\#C_1V(N)C_2$ sequences

Segment Class: $\{C_1, C_2\}$ = coronal plosives, /ɭ/ and /ř/ (= [ʂ])

Table B-5 Dhivehi: Observed values for each $C_1 \ldots C_2$ pair (n = 106)

$C_1 \backslash C_2$	t	d	^{n}d	ʈ	ɖ	^{n}ɖ	ř [ʂ]	ɭ
t	12	4	-	3	7	5	5	12
d	11	10	-	1	2	3	12	15
^{n}d	-	-	-	-	-	-	-	-
ʈ	-	-	-	1	-	-	-	-
ɖ	-	-	-	-	3	-	-	-
^{n}ɖ	-	-	-	-	-	-	-	-
ř [ʂ]	-	-	-	-	-	-	-	-
ɭ	-	-	-	-	-	-	-	-

Table B-6 Dhivehi: Observed (O) and Expected (E) values with O/E ratios; laryngeal distinctions and pre-nasalization ignored (n = 106)

$C_1 \backslash C_2$		t, d, ^{n}d	ʈ, ɖ, ^{n}ɖ	ř [ʂ], ɭ
t, d, ^{n}d	O	37	21	44
	E	35.6	24.1	42.3
	O/E	1.04	0.87	1.04
ʈ, ɖ, ^{n}ɖ	O	0	4	0
	E	1.4	0.9	1.7
	O/E	0.00	4.24	0.00

Gadaba, Konekor

Classification: Dravidian, Central

Descriptive Source(s): Bhaskararao (1980; 1998)

Consonant Phonemes:

p	t	ʈ	ʧ	k
b	d	ɖ	ʤ	g
	s			
m	n	ɳ		ŋ
	r			
	l			
v			j	

Lexical Data Source: Bhaskararao (1980)

Search Domain: headwords containing $\#C_1V(N)C_2$ sequences

Segment Class: $\{C_1, C_2\}$ = coronal plosives and retroflex sonorants

Table B-7 Gadaba: Observed values for each $C_1 \ldots C_2$ pair (n = 38)

$C_1 \backslash C_2$	t	d	ʈ	ɖ	ɳ
t	1	-	10	8	8
d	-	3	2	5	-
ʈ	-	-	1	-	-
ɖ	-	-	-	-	-
ɳ	-	-	-	-	-

Table B-8 Gadaba: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner (n = 38)

$C_1 \backslash C_2$		t, d	ʈ, ɖ	ɳ
t, d	O	4	25	8
	E	3.9	25.3	7.8
	O/E	1.03	0.99	1.03
ʈ, ɖ	O	0	1	0
	E	0.1	0.7	0.2
	O/E	0.00	1.46	0.00

Gondi, Adilabad

Classification:	Dravidian, South-Central
Descriptive Source(s):	Subrahmanyam (1968); Lincoln (1969)

Consonant Phonemes:

p	t	ʈ	ʧ	k	
(p^h)	(t^h)	($ʈ^h$)	($ʧ^h$)	(k^h)	
b	d	ɖ	ʤ	g	
(b^h)	(d^h)	($ɖ^h$)	($ʤ^h$)	(g^h)	
	s		(ʃ)		h
m	n			ŋ	
	r	ɽ			
	l				
v			j		

Lexical Data Source:	Penny et. al. (2005)
Search Domain:	headwords containing $\#C_1V(N)C_2$ sequences
Segment Class:	$\{C_1, C_2\}$ = coronal plosives and retroflex sonorants

Table B-9 Gondi: Observed values for each $C_1 \ldots C_2$ pair (n = 92)

$C_1 \backslash C_2$	t	t^h	d	d^h	ʈ	$ʈ^h$	ɖ	$ɖ^h$	ɽ
t	7	-	4	-	2	-	7	-	13
t^h	-	-	-	-	-	-	-	-	2
d	3	-	7	-	1	-	7	-	5
d^h	-	-	1	-	-	-	1	-	1
ʈ	-	-	-	-	6	-	7	-	-
$ʈ^h$	-	-	-	-	-	-	-	-	-
ɖ	-	-	-	-	1	-	14	-	1
$ɖ^h$	-	-	-	-	1	-	1	-	-
ɽ	-	-	-	-	-	-	-	-	-

Table B-10 Gondi: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner (n = 92)

$C_1 \backslash C_2$		t, t^h, d, d^h	ʈ, $ʈ^h$, ɖ, $ɖ^h$	ɽ
t, t^h, d, d^h	O	22	18	21
	E	14.6	31.8	14.6
	O/E	1.51	0.57	1.44
ʈ, $ʈ^h$, ɖ, $ɖ^h$	O	0	30	1
	E	7.4	16.2	7.4
	O/E	0.00	1.85	0.13

Gorum (Parengi)

Classification:	Munda, South					
Descriptive Source(s):	Anderson & Rau (2008)					
Consonant Phonemes:	p	t	(ʈ)		k	ʔ
	b	(d)	ɖ		g	
		s				
		z				
	m	n		(ɲ)	ŋ	
		r	ɽ			
		l				
				j		
Lexical Data Source:	Donegan & Stampe (2004)					
Search Domain:	headwords containing $\#C_1V(N)C_2$ sequences					
Segment Class:	$\{C_1, C_2\}$ = coronal plosives and retroflex sonorants					

Table B-11 Gorum: Observed values for each $C_1 \ldots C_2$ pair (n = 55)

C_1\ C_2	t	d	ʈ	ɖ	ɽ
t	9	-	1	8	5
d	-	1	-	1	-
ʈ	-	-	4	2	-
ɖ	1	-	1	16	6
ɽ	-	-	-	-	-

Table B-12 Gorum: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner (n = 55)

C_1\ C_2		t, d	ʈ, ɖ	ɽ
t, d	O	10	10	5
	E	5.0	15.0	5.0
	O/E	2.00	0.67	1.00
ʈ, ɖ	O	1	23	6
	E	6.0	18.0	6.0
	O/E	0.17	1.28	1.00

Gtaʔ (Gataʔ, Didayi)

Classification:	Munda, South
Descriptive Source(s):	Anderson (2008a)
Consonant Phonemes:	p t ʈ ʧ k ʔ b (d) ɖ ʤ g s h m n ŋ r ɽ l
Lexical Data Source:	Donegan & Stampe (2004)
Search Domain:	headwords containing $\#C_1V(N)C_2$ sequences
Segment Class:	$\{C_1, C_2\}$ = coronal plosives and retroflex sonorants

Table B-13 Gtaʔ: Observed values for each $C_1 \ldots C_2$ pair (n = 25)

$C_1 \backslash C_2$	t	d	ʈ	ɖ	ɽ
t	2	-	-	2	1
d	-	3	-	-	1
ʈ	-	-	3	-	-
ɖ	2	-	-	11	-
ɽ	-	-	-	-	-

Table B-14 Gtaʔ: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner (n = 25)

$C_1 \backslash C_2$		t, d	ʈ, ɖ	ɽ
t, d	O	5	2	2
	E	2.5	5.8	0.7
	O/E	1.98	0.35	2.78
ʈ, ɖ	O	2	14	0
	E	4.5	10.2	1.3
	O/E	0.45	1.37	0.00

Gutob (Gadaba)

Classification:	Munda, South					
Descriptive Source(s):	Griffiths (2008); Rajan & Rajan (2001)					
Consonant Phonemes:	p	t	(ʈ)		k	ʔ
	b	(d)	ɖ	ʤ	g	
		s				(h)
		z				
	m	n		ɲ	ŋ	
		r	(ɽ)			
		l				
	(w)			(j)		
Lexical Data Source:	Donegan & Stampe (2004)					
Search Domain:	headwords containing $\#C_1V(N)C_2$ sequences					
Segment Class:	$\{C_1, C_2\}$ = coronal plosives and retroflex sonorants					

Table B-15 Gutob: Observed values for each $C_1 \ldots C_2$ pair (n = 32)

$C_1 \backslash C_2$	t	d	ʈ	ɖ	ɽ
t	3	1	-	2	
d	1	3	-	-	2
ʈ	-	1	2	4	-
ɖ	1	-	2	7	3
ɽ	-	-	-	-	-

Table B-16 Gutob: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner (n = 32)

$C_1 \backslash C_2$		t, d	ʈ, ɖ	ɽ
t, d	O	8	2	2
	E	3.8	6.4	1.9
	O/E	2.13	0.31	1.07
ʈ, ɖ	O	2	15	3
	E	6.3	10.6	3.1
	O/E	0.32	1.41	0.96

Hindi

Classification:	Indo-Aryan, Central					
Descriptive Source(s):	Kaye (1997); Ohala (1983; 1994); Shapiro (2003)					
Consonant Phonemes:	p	t	ʈ	ʧ	k	
	pʰ	tʰ	ʈʰ	ʧʰ	kʰ	
	b	d	ɖ	ʤ	g	
	bʰ	dʰ	ɖʰ	ʤʰ	gʰ	
	(f)	s (z)		ʃ		h
	m	n				
		r	ɽ			
			ɽʰ			
		l				
	w			j		
Lexical Data Source:	McGregor (1993)					
Search Domain:	headwords containing #C_1V(N)C_2 sequences					
Segment Class:	{C_1, C_2} = coronal plosives and retroflex sonorants					

Table B-17 Hindi: Observed values for each C_1…C_2 pair (n = 777)

C_1\ C_2	t	tʰ	d	dʰ	ʈ	ʈʰ	ɖ	ɖʰ	ɽ	ɽʰ
t	84	7	31	3	4	-	4	-	55	1
tʰ	5	11	-	-	-	-	-	-	4	-
d	80	1	48	20	2	5	11	-	24	14
dʰ	34	1	1	44	2	-	3	-	26	-
ʈ	-	-	-	-	59	2	8	-	9	4
ʈʰ	-	-	-	-	10	35	9	4	2	3
ɖ	-	-	-	-	9	9	9	1	14	12
ɖʰ	-	-	-	-	7	6	2	24	6	8
ɽ	-	-	-	-	-	-	-	-	-	-
ɽʰ	-	-	-	-	-	-	-	-	-	-

Table B-18 Hindi: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner (n = 777)

C_1\ C_2		t, tʰ, d, dʰ	ʈ, ʈʰ, ɖ, ɖʰ	ɽ, ɽʰ
t, tʰ, d, dʰ	O	370	31	124
	E	250.0	152.0	123.0
	O/E	1.48	0.20	1.01
ʈ, ʈʰ, ɖ, ɖʰ	O	0	194	58
	E	120.0	73.0	59.0
	O/E	0.00	2.66	0.98

Ho

Classification: Munda, North

Descriptive Source(s): Anderson, Osada & Harrison (2008)

Consonant Phonemes:

p	t	ʈ	ʧ	k	ʔ
b	d	ɖ	ʤ	g	
	s				h
m	n	(ɳ)	(ɲ)	(ŋ)	
	r	ɽ			
	l				
w				j	

Lexical Data Source: Deeney (1978)

Search Domain: headwords containing $\#C_1V(N)C_2$ sequences[1]

Segment Class: $\{C_1, C_2\}$ = coronal plosives and retroflex sonorants

Table B-19 Ho: Observed values for each $C_1 \ldots C_2$ pair (n = 178)

$C_1 \backslash C_2$	t	d	ʈ	ɖ	ɽ
t	13	1	1	15	20
d	5	14	6	21	18
ʈ	-	-	24	8	5
ɖ	-	-	5	22	-
ɽ	-	-	-	-	-

Table B-20 Ho: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner (n = 178)

$C_1 \backslash C_2$		t, d	ʈ, ɖ	ɽ
t, d	O	33	43	38
	E	21.1	65.3	27.5
	O/E	1.56	0.66	1.38
ʈ, ɖ	O	0	59	5
	E	11.9	36.7	15.5
	O/E	0.00	1.61	0.32

[1] The count does not include lexical entries marked by an asterisk (*) in Deeney (1978). Such entries represent compounds derived from a preceeding headword (p. vi). Words were also excluded from the count if C_1 and C_2 are separated by a hyphen in Deeney's transcription. These cases involve reduplication or some other form of morphological complexity (e.g., dū-dū 'of much smoke, to arise').

Indus Kohistani (Maiyã)

Classification: Indo-Aryan, Northwest, Dardic

Descriptive Source(s): Zoller (2005); cf. Hallberg and Hallberg (1999)

Consonant Phonemes:

p	t	ʈ		k	(q)	
p^h	t^h	ʈ^h		k^h		
b	d	ɖ		g		
b^h	d^h	ɖ^h		g^h		
	ʦ	ʈʂ	ʧ			
	ʦ^h	ʈʂ^h	ʧ^h			
(f)	s	ʂ	ʃ		(x)	h
	z	ʐ	ʒ		(ɣ)	
	z^h	ʐ^h	ʒ^h			
m	n					
	l					
	r	ɽ				
w			j			

Lexical Data Source: Zoller (2005)

Search Domain 1: headwords containing $\#C_1V(N)C_2$ sequences

Search Domain 2: roots containing $\#C_1V(N)C_2$ sequences

Segment Class: $\{C_1, C_2\}$ = coronal obstruents and retroflex sonorants

Table B-21 Indus Kohistani: Observed values for each $C_1 \ldots C_2$ pair in dictionary *headwords* containing $\#C_1V(N)C_2$ sequences (n = 627)

C_1\C_2	t	tʰ	d	dʰ	ʈ	ʈʰ	ɖ	ɖʰ	ts	tsʰ	tʃ	tʃʰ	ʈʂ	ʈʂʰ	s	z	zʰ	ʃ	ʒ	ʒʰ	ʂ	ʐ	ʐʰ	ɽ
t	17	3	7	-	-	-	1	-	-	-	-	-	-	1	12	8	-	2	3	-	-	-	-	6
tʰ	9	-	-	-	-	1	-	-	-	-	-	-	2	-	-	-	-	2	-	-	2	-	-	3
d	-	1	15	-	-	2	-	-	-	-	-	-	-	2	12	8	-	13	1	-	3	-	-	6
dʰ	4	1	1	1	-	-	-	-	-	-	-	-	-	-	3	2	-	-	1	-	2	-	-	9
ʈ	-	-	-	-	11	2	4	-	-	1	-	-	-	-	4	-	-	-	-	-	-	-	-	1
ʈʰ	-	-	-	-	2	4	1	-	-	-	-	-	-	-	3	-	-	-	-	-	1	-	-	-
ɖ	-	-	-	-	1	2	19	-	-	-	-	-	-	-	1	2	-	1	-	-	-	-	-	3
ɖʰ	-	-	-	-	-	-	12	-	1	-	-	-	-	-	3	-	-	-	-	-	-	-	-	1
ts	-	-	6	-	7	1	5	-	2	2	-	-	-	-	3	2	-	1	-	-	-	-	-	-
tsʰ	3	-	-	-	3	1	1	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-
tʃ	11	1	6	-	7	10	2	-	-	-	6	-	-	-	1	-	-	2	1	-	-	-	-	-
tʃʰ	1	1	-	-	3	1	-	-	-	-	-	-	-	-	-	-	-	1	3	-	-	-	-	-
ʈʂ	-	-	-	-	-	2	-	-	-	-	-	-	-	1	2	-	-	-	-	-	10	-	-	-
ʈʂʰ	2	1	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	5	-	-
s	23	13	17	-	2	3	15	-	-	1	4	-	2	-	8	7	-	1	1	-	1	-	-	-
z	6	2	7	-	2	1	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-
zʰ	2	-	-	-	1	-	4	-	-	-	-	-	-	-	-	7	-	-	-	-	-	-	-	-
ʃ	7	3	9	-	3	2	9	-	-	-	1	-	-	-	-	3	-	15	2	-	-	-	-	-
ʒ	1	-	7	-	-	1	1	-	-	-	-	-	-	-	3	-	-	1	6	-	-	-	-	-
ʒʰ	-	-	-	-	1	1	-	-	-	-	-	-	-	-	-	1	-	1	-	-	-	-	-	-
ʂ	-	-	-	-	1	2	-	-	-	-	-	-	-	-	1	5	-	3	-	-	15	-	-	1
ʐ	-	-	-	-	-	-	2	-	-	-	-	-	1	-	-	-	-	-	-	-	2	1	-	-
ʐʰ	-	-	-	1	3	3	-	-	-	-	-	-	1	-	1	1	-	-	-	-	7	-	-	-
ɽ	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table B-22 Indus Kohistani: Observed (O) and Expected (E) values with O/E ratios for pairs of coronal plosives and retroflex sonorants in *headwords* containing $\#C_1V(N)C_2$ sequences; segments classified by place and manner (n = 150)

C_1\ C_2		t, tʰ, d, dʰ	ʈ, ʈʰ, ɖ, ɖʰ	ɽ
t, tʰ, d, dʰ	O	59	4	24
	E	34.2	36.0	16.8
	O/E	1.72	0.11	1.43
ʈ, ʈʰ, ɖ, ɖʰ	O	0	58	5
	E	24.8	26.0	12.2
	O/E	0.00	2.23	0.41

Table B-23 Indus Kohistani: Observed (O) and Expected (E) values with O/E ratios for pairs of coronal obstruents in ***headwords*** containing #$C_1V(N)C_2$ sequences; segments classified by place and manner; linear order of segments collapsed (n = 597)

C_1\ C_2		t, t^{h}, d, d^{h}	ʈ, ʈ^{h}, ɖ, ɖ^{h}	ʦ, ʦ^{h}	ʧ, ʧ^{h} ʒ-, ʒ^{h}-	ʈʂ, ʈʂ^{h}	s, z, z^{h}	ʃ, -ʒ, -ʒ^{h}	ʂ, ʐ, ʐ^{h}
	O	59	4	9	28	8	115	41	8
t, t^{h}, d, d^{h}	E	45.0	62.3	14.0	28.3	10.2	67.2	31.6	27.5
	O/E	1.31	0.06	0.64	0.99	0.78	1.71	1.30	0.29
	O		58	20	27	2	41	15	12
ʈ, ʈ^{h}, ɖ, ɖ^{h}	E		20.4	11.5	23.4	7.9	49.3	22.3	19.7
	O/E		2.85	1.74	1.16	0.25	0.83	0.67	0.61
	O			4	0	0	8	1	0
ʦ, ʦ^{h}	E			0.5	1.7	1.0	8.5	4.6	3.8
	O/E			8.75	0.00	0.00	0.94	0.22	0.00
	O				6	0	9	16	0
ʧ, ʧ^{h} ʒ-, ʒ^{h}-	E				1.5	1.9	16.9	9.3	7.6
	O/E				4.02	0.00	0.53	1.73	0.00
	O					2	4	0	17
ʈʂ, ʈʂ^{h}	E					0.4	6.7	3.4	2.9
	O/E					4.52	0.59	0.00	5.88
	O						24	5	9
s, z, z^{h}	E						23.7	23.2	19.8
	O/E						1.01	0.22	0.45
	O							17	3
ʃ, -ʒ, -ʒ^{h}	E							5.5	9.5
	O/E							3.08	0.31
	O								25
ʂ, ʐ, ʐ^{h}	E								4.1
	O/E								6.09

Table B-24 Indus Kohistani: Observed (O) and Expected (E) values with O/E ratios for pairs of coronal obstruents in *headwords* containing #$C_1V(N)C_2$ sequences; segments classified by place and sibilant/non-sibilant manner; linear order collapsed (n = 597)

C₁\ C₂		t, tʰ, d, dʰ	ʈ, ʈʰ, ɖ, ɖʰ	ts, tsʰ, s, z, zʰ	tʃ, tʃʰ, ʃ, ʒ, ʒʰ	ʈʂ, ʈʂʰ, ʂ, ʐ, ʐʰ
t, tʰ, d, dʰ	O	59	4	124	69	16
	E	45.0	62.3	81.3	59.9	37.7
	O/E	1.31	0.06	1.53	1.15	0.42
ʈ, ʈʰ, ɖ, ɖʰ	O		58	61	42	14
	E		20.4	60.7	45.7	27.6
	O/E		2.85	1.00	0.92	0.51
ts, tsʰ, s, z, zʰ	O			36	15	13
	E			32.7	46.4	31.3
	O/E			1.10	0.32	0.42
tʃ, tʃʰ, ʃ, ʒ, ʒʰ	O				39	3
	E				16.3	22.5
	O/E				2.40	0.13
ʈʂ, ʈʂʰ, ʂ, ʐ, ʐʰ	O					44
	E					7.4
	O/E					5.92

Table B-25 Indus Kohistani: Observed (O) and Expected (E) values with O/E ratios for pairs of coronal obstruents in *headwords* containing #$C_1V(N)C_2$ sequences; segments classified as retroflex/non-retroflex and sibilant/non-sibilant; linear order of segments collapsed (n = 597)

C₁\ C₂		t, tʰ, d, dʰ	ʈ, ʈʰ, ɖ, ɖʰ	ts, tsʰ, s, z, zʰ tʃ, tʃʰ, ʃ, ʒ, ʒʰ	ʈʂ, ʈʂʰ, ʂ, ʐ, ʐʰ
t, tʰ, d, dʰ	O	59	4	193	16
	E	45.0	62.3	141.1	37.7
	O/E	1.31	0.06	1.37	0.42
ʈ, ʈʰ, ɖ, ɖʰ	O		58	103	14
	E		20.4	106.4	27.6
	O/E		2.85	0.97	0.51
ts, tsʰ, s, z, zʰ tʃ, tʃʰ, ʃ, ʒ, ʒʰ	O			90	16
	E			95.3	53.8
	O/E			0.94	0.30
ʈʂ, ʈʂʰ, ʂ, ʐ, ʐʰ	O				44
	E				7.4
	O/E				5.92

Table B-26 Indus Kohistani: Observed values for each $C_1 \ldots C_2$ pair in lexical *roots* containing $\#C_1V(N)C_2$ sequences (n = 324)

$C_1 \backslash C_2$	t	tʰ	d	dʰ	ʈ	ʈʰ	ɖ	ɖʰ	ts	tsʰ	tʃ	tʃʰ	ʈʂ	ʈʂʰ	s	z	zʰ	ʃ	ʒ	ʒʰ	ʂ	ʐ	ʐʰ	ɽ
t	7	2	6	-	-	-	1	-	-	-	-	-	-	1	12	5	-	-	3	-	-	-	-	5
tʰ	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	3
d	-	1	10	-	-	1	-	-	-	-	-	-	-	1	5	2	-	5	-	-	2	-	-	2
dʰ	2	1	1	-	-	-	-	-	-	-	-	-	-	-	2	1	-	-	1	-	2	-	-	7
ʈ	-	-	-	-	3	2	2	-	-	1	-	-	-	-	2	-	-	-	-	-	-	-	-	1
ʈʰ	-	-	-	-	-	4	1	-	-	-	-	-	-	-	2	-	-	-	-	-	1	-	-	-
ɖ	-	-	-	-	-	1	5	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	1
ɖʰ	-	-	-	-	-	-	3	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	1
ts	-	-	3	-	5	-	2	-	-	1	-	-	-	-	1	2	-	-	-	-	-	-	-	-
tsʰ	2	-	-	-	-	-	1	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-
tʃ	4	-	2	-	5	7	2	-	-	-	3	-	-	-	1	-	-	1	1	-	-	-	-	-
tʃʰ	1	1	-	-	3	1	-	-	-	-	-	-	-	-	-	-	-	1	2	-	-	-	-	-
ʈʂ	-	-	-	-	-	2	-	-	-	-	-	-	-	1	-	-	-	-	-	-	6	-	-	-
ʈʂʰ	2	1	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	3	-	-
s	9	8	12	-	1	2	5	-	-	-	3	-	1	-	2	4	-	1	1	-	-	-	-	-
z	2	2	3	-	-	1	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-
zʰ	1	-	-	-	1	-	2	-	-	-	-	-	-	-	-	3	-	-	-	-	-	-	-	-
ʃ	6	3	4	-	2	1	5	-	-	-	1	-	-	-	-	1	-	7	2	-	-	-	-	-
ʒ	1	-	4	-	-	1	1	-	-	-	-	-	-	-	1	-	-	-	3	-	-	-	-	-
ʒʰ	-	-	-	-	1	1	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-
ʂ	-	-	-	-	1	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	7	-	-	1
ʐ	-	-	-	-	-	-	2	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-
ʐʰ	-	-	-	-	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-
ɽ	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table B-27 Indus Kohistani: Observed (O) and Expected (E) values with O/E ratios for pairs of coronal obstruents in *roots* containing #$C_1V(N)C_2$ sequences; segments classified by place and manner; linear order of segments collapsed (n = 303)

C_1\ C_2		t, t^h, d, d^h	ʈ, ʈ^h, ɖ, ɖ^h	ʦ, ʦ^h	ʧ, ʧ^h ʒ-, ʒ^h-	ʈʂ, ʈʂ^h	s, z, z^h	ʃ, -ʒ, -ʒ^h	ʂ, ʐ, ʐ^h
t, t^h, d, d^h	O	31	2	5	13	5	64	22	5
	E	25.6	30.7	6.6	18.3	6.9	34.9	17.8	11.7
	O/E	1.21	0.07	0.76	0.71	0.73	1.83	1.24	0.43
ʈ, ʈ^h, ɖ, ɖ^h	O		21	9	22	2	19	8	8
	E		8.1	5.1	14.0	4.9	22.7	11.4	7.0
	O/E		2.59	1.78	1.58	0.41	0.84	0.70	1.15
ʦ, ʦ^h	O			1	0	0	4	0	0
	E			0.1	0.7	0.5	3.5	1.9	1.5
	O/E			8.42	0.00	0.00	1.14	0.00	0.00
ʧ, ʧ^h ʒ-, ʒ^h-	O				3	0	6	9	0
	E				1.1	1.3	9.9	5.3	4.3
	O/E				2.65	0.00	0.61	1.71	0.00
ʈʂ, ʈʂ^h	O					2	1	0	10
	E					0.3	4.0	2.1	1.6
	O/E					6.31	0.25	0.00	6.23
s, z, z^h	O						10	3	0
	E						11.2	11.5	8.1
	O/E						0.90	0.26	0.00
ʃ, -ʒ, -ʒ^h	O							9	0
	E							3.0	4.1
	O/E							3.04	0.00
ʂ, ʐ, ʐ^h	O								9
	E								1.3
	O/E								6.68

Table B-28 Indus Kohistani: Observed (O) and Expected (E) values with O/E ratios for pairs of coronal obstruents in *roots* containing #C_1V(N)C_2 sequences; segments classified by place and sibilant/non-sibilant manner; linear order collapsed (n = 303)

C_1\ C_2		t, tʰ, d, dʰ	ʈ, ʈʰ, ɖ, ɖʰ	ʦ, ʦʰ, s, z, zʰ	ʧ, ʧʰ, ʃ, ʒ, ʒʰ	ʈʂ, ʈʂʰ, ʂ, ʐ, ʐʰ
t, tʰ, d, dʰ	O	31	2	69	35	10
	E	25.6	30.7	41.5	36.0	18.6
	O/E	1.21	0.07	1.66	0.97	0.54
ʈ, ʈʰ, ɖ, ɖʰ	O		21	28	30	10
	E		8.1	27.8	25.4	11.9
	O/E		2.59	1.01	1.18	0.84
ʦ, ʦʰ, s, z, zʰ	O			15	9	1
	E			14.8	24.0	14.1
	O/E			1.01	0.37	0.07
ʧ, ʧʰ, ʃ, ʒ, ʒʰ	O				21	0
	E				9.4	11.8
	O/E				2.24	0.00
ʈʂ, ʈʂʰ, ʂ, ʐ, ʐʰ	O					21
	E					3.3
	O/E					6.43

Table B-29 Indus Kohistani: Observed (O) and Expected (E) values with O/E ratios for pairs of coronal obstruents in *roots* containing #C_1V(N)C_2 sequences; segments classified as retroflex/non-retroflex and sibilant/non-sibilant; linear order of segments collapsed (n = 303)

C_1\ C_2		t, tʰ, d, dʰ	ʈ, ʈʰ, ɖ, ɖʰ	ʦ, ʦʰ, s, z, zʰ ʧ, ʧʰ, ʃ, ʒ, ʒʰ	ʈʂ, ʈʂʰ, ʂ, ʐ, ʐʰ
t, tʰ, d, dʰ	O	31	2	104	10
	E	25.6	30.7	77.5	18.6
	O/E	1.21	0.07	1.34	0.54
ʈ, ʈʰ, ɖ, ɖʰ	O		21	58	10
	E		8.1	53.2	11.9
	O/E		2.59	1.09	0.84
ʦ, ʦʰ, s, z, zʰ ʧ, ʧʰ, ʃ, ʒ, ʒʰ	O			45	1
	E			48.2	25.9
	O/E			0.93	0.04
ʈʂ, ʈʂʰ, ʂ, ʐ, ʐʰ	O				21
	E				3.3
	O/E				6.43

Juang

Classification:	Munda, South					
Descriptive Source(s):	Mathew & Mathew (2003); Matson (1964); Patnaik (2008)					
Consonant Phonemes:[2]	p	t	ʈ	ʧ	k	
	(p^h)	(t^h)	(ʈ^h)	(ʧ^h)	(k^h)	
	b	d	ɖ	ʤ	g	
	(b^h)	(d^h)	(ɖ^h)	(ʤ^h)	(g^h)	
		s				(h)
	m	n	ɳ	ɲ	ŋ	
		r	(ɽ)			
			(ɽ^h)			
		l	ɭ			
				j		
Lexical Data Source:	Donegan & Stampe (2004)					
Search Domain:	headwords containing #C_1V(N)C_2 sequences					
Segment Class:	{C_1, C_2} = coronal plosives and retroflex sonorants					

Table B-30 Juang: Observed values for each C_1…C_2 pair (n = 68)

C_1\ C_2	t	t^h	d	d^h	ʈ	ʈ^h	ɖ	ɖ^h	ɳ	ɭ	ɽ	ɽ^h
t	9	-	-		3	-	1	-	1	6	4	-
t^h	-	-	-	-	-	-	-	-	-	1	-	-
d	2	-	1	1	-	-	-	-	1	2	6	-
d^h	1	-	-	-	-	-	-	-	2	-	-	-
ʈ	1	-	-	-	3	-	-	-	2	2	-	-
ʈ^h	-	-	-	-	-	-	-	-	-	1	-	-
ɖ	-	-	-	-	1	-	8	-	1	4	2	1
ɖ^h	-	-	-	-	-	-	-	1	-	-	-	-
ɳ	-	-	-	-	-	-	-	-	-	-	-	-
ɭ	-	-	-	-	-	-	-	-	-	-	-	-
ɽ	-	-	-	-	-	-	-	-	-	-	-	-
ɽ^h	-	-	-	-	-	-	-	-	-	-	-	-

[2] /h/, /ɽ/ and the aspirated consonants are not listed as phonemes in most of the descriptive sources but are included in Donegan & Stampe (2004), which is the source of lexical data for the present study. The aspirates and /h/ are probably the result of Indo-Aryan loanwords. Most sources treat [ɽ] as an allophone of /ɖ/ or /ɭ/ in native vocabulary.

Table B-31 Juang: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner (n = 68)

C_1\ C_2		t, t^h, d, d^h	ʈ, ʈ^h, ɖ, ɖ^h	ɳ, ɭ, ɽ, ɽ^h
t, t^h, d, d^h	O	14	4	23
	E	9.0	10.3	21.7
	O/E	1.55	0.39	1.06
ʈ, ʈ^h, ɖ, ɖ^h	O	1	13	13
	E	6.0	6.8	14.3
	O/E	0.17	1.93	0.91

Kalami (Kalam Kohistani)

Classification:	Indo-Aryan, Northwest, Dardic						
Descriptive Source(s):	Baart (1997)						
Consonant Phonemes:[3]	p	t	ʈ		k	(q)	
	p^h	t^h	ʈ^h		k^h		
	b	d	ɖ		g		
			(ɖ^h)				
		ʦ	ʈʂ	ʧ			
		ʦ^h	ʈʂ^h	ʧ^h			
				ʤ			
				(ʤ^h)			
	(f)	s	ʂ	ʃ	(x)		h
		z	(ʐ)		(ɣ)		
	m	n	ɳ		ŋ		
		r	ɽ				
		l					
		ɫ					
	w			j			
Lexical Data Source:	Joan Baart (personal communication, Toolbox database)						
Search Domain:	headwords containing #C_1V(N)C_2 sequences						
Segment Class:	{C_1, C_2} = coronal obstruents (plosives, affricates, fricatives)						

[3] [ɖ^h], [ʤ^h] and [ʐ] are not listed as phonemes in Baart (1997). They are included here because at least one instance of each was found in the lexical data source. /q, f, z, x, ɣ/ occur mainly in loanwords.

Table B-32 Kalami: Observed values for each $C_1...C_2$ pair (n = 468)

C_1\\C_2	t	t^h	d	ʈ	ʈ^h	ɖ	ɖ^h	ts	ts^h	tʃ	tʃ^h	dʒ	dʒ^h	ʈʂ	ʈʂ^h	s	z	ʃ	ʂ	ʐ
t	12	6	4	-	-	-	-	-	-	-	-	7	-	2	-	14	6	5	4	-
t^h	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	1	1	-
d	6	-	8	-	-	-	-	-	-	2	3	4	-	-	-	11	1	8	-	-
ʈ	-	-	-	8	-	-	-	-	-	1	-	-	-	-	-	9	-	3	-	-
ʈ^h	-	-	-	3	-	1	-	-	-	-	-	-	-	-	-	1	-	-	-	-
ɖ	-	1	-	3	2	10	-	1	-	1	-	2	-	-	-	5	6	-	-	-
ɖ^h	-	-	-	1	-	1	-	-	-	2	-	-	-	-	-	1	-	-	-	-
ts	-	-	1	2	-	-	-	2	-	-	-	-	-	-	-	-	1	-	-	-
ts^h	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-
tʃ	8	-	4	29	-	1	-	-	-	12	-	4	-	-	-	6	1	2	4	-
tʃ^h	3	-	3	2	-	-	-	-	-	3	-	2	-	-	-	-	-	-	-	-
dʒ	4	-	8	4	-	2	-	-	-	2	2	10	-	-	-	8	5	3	3	-
dʒ^h	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ʈʂ	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ʈʂ^h	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-
s	23	-	15	6	-	1	-	-	-	9	-	7	-	3	-	2	6	2	-	-
z	5	-	6	-	-	1	-	-	-	-	-	-	-	-	-	-	1	-	-	-
ʃ	7	-	9	17	-	1	-	-	-	-	-	-	-	-	-	-	1	16	-	-
ʂ	2	-	-	-	-	-	-	-	-	1	-	3	-	-	-	-	-	-	7	-
ʐ	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1

Table B-33 Kalami: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner; linear order of segments collapsed (n = 468)

C_1\C_2		t, tʰ, d	ʈ, ʈʰ, ɖ, ɖʰ	ʦ, ʦʰ	ʧ, ʧʰ, ʤ, ʤʰ	ʈʂ, ʈʂʰ	s, z	ʃ	ʂ, ʐ
t, tʰ, d	O	39	1	1	46	2	82	30	7
	E	32.1	40.6	2.8	58.3	2.0	45.9	24.4	8.8
	O/E	1.21	0.02	0.36	0.79	1.01	1.79	1.23	0.80
ʈ, ʈʰ, ɖ, ɖʰ	O		29	3	45	0	30	21	0
	E		12.7	1.8	38.2	1.2	29.4	15.8	5.5
	O/E		2.28	1.64	1.18	0.00	1.02	1.33	0.00
ʦ, ʦʰ	O			2	0	0	2	0	0
	E			0.0	2.0	0.1	1.9	0.9	0.4
	O/E			44.57	0.00	0.00	1.08	0.00	0.00
ʧ, ʧʰ, ʤ, ʤʰ	O				35	1	36	5	11
	E				22.7	2.1	39.8	20.1	8.1
	O/E				1.54	0.48	0.90	0.25	1.35
ʈʂ, ʈʂʰ	O					1	3	0	0
	E					0.0	1.5	0.8	0.3
	O/E					39.00	2.02	0.00	0.00
s, z	O						9	3	0
	E						16.2	16.9	6.3
	O/E						0.56	0.18	0.00
ʃ	O							16	0
	E							4.4	3.4
	O/E							3.67	0.00
ʂ, ʐ	O								8
	E								0.6
	O/E								13.37

Table B-34 Kalami: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and sibilant/non-sibilant manner; linear order collapsed (n = 468)

C$_1$\ C$_2$		t, t^h, d	ʈ, ʈ^h, ɖ, ɖ^h	ʦ, ʦ^h, s, z	ʧ, ʧ^h, ʤ, ʤ^h, ʃ	ʈʂ, ʈʂ^h, ʂ, ʐ
t, t^h, d	O	39	1	83	76	9
	E	32.1	40.6	48.7	82.6	10.8
	O/E	1.21	0.02	1.71	0.92	0.84
ʈ, ʈ^h, ɖ, ɖ^h	O		29	33	66	0
	E		12.7	31.2	54.0	6.7
	O/E		2.28	1.06	1.22	0.00
ʦ, ʦ^h, s, z	O			13	39	3
	E			18.1	59.7	8.3
	O/E			0.72	0.65	0.36
ʧ, ʧ^h, ʤ, ʤ^h, ʃ	O				56	12
	E				47.1	14.4
	O/E				1.19	0.83
ʈʂ, ʈʂ^h, ʂ, ʐ	O					9
	E					0.9
	O/E					10.13

Table B-35 Kalami: Observed (O) and Expected (E) values with O/E ratios; segments classified as retroflex vs. non-retroflex and sibilant vs. non-sibilant; linear order of segments collapsed (n = 468)

C$_1$\ C$_2$		t, t^h, d	ʈ, ʈ^h, ɖ, ɖ^h	ʦ, ʦ^h, s, z ʧ, ʧ^h, ʤ, ʤ^h, ʃ	ʈʂ, ʈʂ^h, ʂ, ʐ
t, t^h, d	O	39	1	159	9
	E	32.1	40.6	131.3	10.8
	O/E	1.21	0.02	1.21	0.84
ʈ, ʈ^h, ɖ, ɖ^h	O		29	99	0
	E		12.7	85.2	6.7
	O/E		2.28	1.16	0.00
ʦ, ʦ^h, s, z ʧ, ʧ^h, ʤ, ʤ^h, ʃ	O			108	15
	E			124.9	22.7
	O/E			0.86	0.66
ʈʂ, ʈʂ^h, ʂ, ʐ	O				9
	E				0.9
	O/E				10.13

Kalasha

Classification: Indo-Aryan, Northwest, Dardic

Descriptive Source(s): Heegård & Mørch (2004), Trail & Cooper (1999)

Consonant Phonemes:[4]	p	t		ʈ		k	
	pʰ	tʰ		ʈʰ		k	
	b	d		ɖ		g	
	bʰ	dʰ		ɖʰ		gʰ	
		ʦ		ʈʂ	ʧ		
		ʦʰ		ʈʂʰ	ʧʰ		
		ʣ		ɖʐ	ʤ		
					ʤʰ		
		s		ʂ	ʃ		h
		z		ʐ	ʒ		
	m	n		(ɳ)		ŋ	
		ɬ	l				
			r	(ɽ)			
	w				j		

Lexical Data Source: Ron Trail & Greg Cooper (personal communication, Toolbox database, cf. Trail & Cooper, 1999)

Search Domain 1: headwords containing $C_1 \ldots C_2$ sequences

Search Domain 2: roots containing $C_1 \ldots C_2$ sequences

Segment Class: $\{C_1, C_2\}$ = coronal obstruents (plosives, affricates, fricatives)

[4] /ɽ/ and /ɳ/ are not distinguished in Trail & Cooper's Toolbox database or in their (1999) Kalasha dictionary. /ɽ/ occurs only in the Birir dialect. /ɳ/ is listed as a phoneme in Heegård & Mørch (2004).

Table B-36 Kalasha: Observed values for pairs of non-adjacent coronal obstruents co-occurring in dictionary *headwords* (n = 994)

$C_1 \backslash C_2$	t	tʰ	d	dʰ	ʈ	ʈʰ	ɖ	ɖʰ	ʦ	ʦʰ	ʥ	ʧ	ʧʰ	ʤ	ʤʰ	ʈʂ	ʈʂʰ	ɖʐ	s	z	ʃ	ʒ	ʂ	ʐ
t	23	1	12	-	-	-	2	-	2	-	-	10	-	7	-	4	-	-	24	5	11	4	6	-
tʰ	-	1	2	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	8	2	2	-	-	-
d	28	-	27	-	4	-	-	-	2	-	-	5	3	5	-	8	4	-	27	7	14	4	16	3
dʰ	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-
ʈ	2	-	1	-	23	3	2	-	2	-	-	5	-	1	-	1	-	-	3	-	3	-	-	-
ʈʰ	-	-	-	-	4	-	-	-	-	-	-	1	-	-	-	-	-	-	3	-	2	-	1	-
ɖ	-	1	1	-	3	-	24	-	-	-	-	-	1	1	-	-	-	-	3	6	6	-	1	-
ɖʰ	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	5	-	2	-	-	-
ʦ	4	-	2	-	1	-	1	-	4	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
ʦʰ	2	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ʥ	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	-	-	-	-
ʧ	18	-	3	1	22	1	5	-	1	-	-	9	-	1	-	-	-	-	8	1	6	4	11	-
ʧʰ	5	-	3	-	3	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	1	-	-	-
ʤ	17	-	5	-	13	-	-	-	-	-	-	-	-	10	-	-	-	-	16	5	1	-	18	1
ʤʰ	2	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
ʈʂ	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	7	4	1	-	-	-	-	3	-
ʈʂʰ	4	-	3	-	-	-	-	-	-	-	-	1	-	-	-	5	1	-	-	-	-	-	-	-
ɖʐ	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3	1	1	-	-	-	-	-	-
s	38	5	26	1	10	1	2	-	-	-	-	9	-	7	4	-	1	-	36	6	5	3	1	1
z	12	-	9	-	-	2	-	-	-	-	-	7	-	1	-	1	-	-	1	3	1	-	-	-
ʃ	18	5	6	-	8	2	2	-	-	-	-	-	-	5	-	-	-	-	3	-	22	1	-	-
ʒ	6	-	4	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	1	1	4	-	-	-
ʂ	5	-	1	-	10	-	1	-	-	-	-	3	-	2	-	-	-	1	2	-	-	-	31	1
ʐ	1	-	-	-	-	-	-	-	-	-	-	-	-	1	-	1	-	-	-	-	-	-	1	1

Table B-37 Kalasha: Observed (O) and Expected (E) values with O/E ratios for pairs of non-adjacent coronal obstruents co-occurring in *headwords*; segments classified by place and manner; linear order of segments collapsed (n = 994)

$C_1 \backslash C_2$		t, t^{h}, d, d^{h}	ʈ, ʈ^{h}, ɖ, ɖ^{h}	ʦ, ʦ^{h}, ʥ	ʧ, ʧ^{h} ʤ, ʤ^{h}	ʈʂ, ʈʂ^{h}, ɖʐ	s, z	ʃ, ʒ	ʂ, ʐ
t, t^{h}, d, d^{h}	O	95	13	13	86	26	164	75	32
	E	90.1	78.8	9.7	90.4	24.3	112.6	55.9	47.2
	O/E	1.05	0.17	1.34	0.95	1.07	1.46	1.34	0.68
ʈ, ʈ^{h}, ɖ, ɖ^{h}	O		59	4	54	1	35	25	13
	E		17.0	4.4	41.0	10.6	49.6	24.4	20.2
	O/E		3.47	0.92	1.32	0.09	0.71	1.02	0.64
ʦ, ʦ^{h}, ʥ	O			5	1	0	3	1	0
	E			0.2	4.5	1.3	6.0	3.0	2.6
	O/E			20.12	0.22	0.00	0.50	0.33	0.00
ʧ, ʧ^{h} ʤ, ʤ^{h}	O				21	1	59	17	36
	E				19.8	12.4	54.9	28.1	25.1
	O/E				1.06	0.08	1.07	0.61	1.43
ʈʂ, ʈʂ^{h}, ɖʐ	O					23	2	0	5
	E					1.6	15.2	7.6	6.3
	O/E					14.04	0.13	0.00	0.79
s, z	O						46	14	4
	E						34.9	35.0	29.9
	O/E						1.32	0.40	0.13
ʃ, ʒ	O							27	0
	E							8.7	14.6
	O/E							3.11	0.00
ʂ, ʐ	O								34
	E								6.0
	O/E								5.68

Table B-38 Kalasha: Observed (O) and Expected (E) values with O/E ratios for pairs of non-adjacent coronal obstruents co-occurring in *headwords*; segments classified by place and sibilant/non-sibilant manner; linear order collapsed (n = 994)

C1\ C2		t, tʰ, d, dʰ	ʈ, ʈʰ, ɖ, ɖʰ	ʦ, ʦʰ, ʣ, s, z	ʧ, ʧʰ, ʤ, ʤʰ, ʃ, ʒ	ʈʂ, ʈʂʰ, ɖʐ, ʂ, ʐ
t, tʰ, d, dʰ	O	95	13	177	161	58
	E	90.1	78.8	122.3	146.3	71.5
	O/E	1.05	0.17	1.45	1.10	0.81
ʈ, ʈʰ, ɖ, ɖʰ	O		59	39	79	14
	E		17.0	54.0	65.5	30.8
	O/E		3.47	0.72	1.21	0.45
ʦ, ʦʰ, ʣ, s, z	O			54	75	6
	E			41.2	97.4	49.1
	O/E			1.31	0.77	0.12
ʧ, ʧʰ, ʤ, ʤʰ, ʃ, ʒ	O				65	37
	E				56.6	59.7
	O/E				1.15	0.62
ʈʂ, ʈʂʰ, ɖʐ, ʂ, ʐ	O					62
	E					13.9
	O/E					4.45

Table B-39 Kalasha: Observed (O) and Expected (E) values with O/E ratios for pairs of non-adjacent coronal obstruents co-occurring in *headwords*; segments classified as retroflex vs. non-retroflex and sibilant vs. non-sibilant; linear order of segments collapsed (n = 994)

C1\ C2		t, tʰ, d, dʰ	ʈ, ʈʰ, ɖ, ɖʰ	ʦ, ʦʰ, ʣ, s, z ʧ, ʧʰ, ʤ, ʤʰ, ʃ, ʒ	ʈʂ, ʈʂʰ, ɖʐ, ʂ, ʐ
t, tʰ, d, dʰ	O	95	13	338	58
	E	90.1	78.8	268.6	71.5
	O/E	1.05	0.17	1.26	0.81
ʈ, ʈʰ, ɖ, ɖʰ	O		59	118	14
	E		17.0	119.4	30.8
	O/E		3.47	0.99	0.45
ʦ, ʦʰ, ʣ, s, z ʧ, ʧʰ, ʤ, ʤʰ, ʃ, ʒ	O			194	43
	E			195.1	108.8
	O/E			0.99	0.40
ʈʂ, ʈʂʰ, ɖʐ, ʂ, ʐ	O				62
	E				13.9
	O/E				4.45

Table B-40 Kalasha: Observed values for pairs of non-adjacent coronal obstruents co-occurring in *roots* (n = 766)

C_1\\C_2	t	tʰ	d	dʰ	ʈ	ʈʰ	ɖ	ɖʰ	ʦ	ʦʰ	ʣ	ʧ	ʧʰ	ʤ	ʤʰ	ʈʂ	ʈʂʰ	ɖʐ	s	z	ʃ	ʒ	ʂ	ʐ
t	17	-	9	-	-	-	1	-	2	-	-	9	-	5	-	3	-	-	18	5	9	3	5	-
tʰ	-	-	2	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	8	2	2	-	-	-
d	24	-	17	-	2	-	-	-	2	-	-	4	2	2	-	7	3	-	24	6	10	3	10	2
dʰ	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-
ʈ	2	-	-	-	14	-	1	-	2	-	-	5	-	1	-	1	-	-	3	-	2	-	-	-
ʈʰ	-	-	-	-	4	-	-	-	-	-	-	1	-	-	-	-	-	-	3	-	2	-	1	-
ɖ	-	-	-	-	2	-	18	-	-	-	-	-	1	1	-	-	-	-	2	5	4	-	-	-
ɖʰ	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	4	-	2	-	-	-
ʦ	3	-	2	-	1	-	1	-	3	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
ʦʰ	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ʣ	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	-	-	-	-
ʧ	15	-	3	-	19	1	4	-	1	-	-	6	-	1	-	-	-	-	4	1	5	3	9	-
ʧʰ	5	-	2	-	2	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	1	-	-	-
ʤ	14	-	3	-	6	-	-	-	-	-	-	-	-	8	-	-	-	-	14	4	2	-	14	1
ʤʰ	2	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
ʈʂ	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	6	1	1	-	-	-	-	3	-
ʈʂʰ	4	-	2	-	-	-	-	-	-	-	-	1	-	-	-	5	1	-	-	-	-	-	-	-
ɖʐ	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	1	-	-	-	-	-	-
s	34	4	24	1	3	1	-	-	-	-	-	7	-	6	3	-	1	-	28	5	3	2	-	1
z	10	-	7	-	-	1	-	-	-	-	-	7	-	1	-	-	-	-	1	1	1	-	-	-
ʃ	15	4	5	-	7	2	2	-	-	-	-	-	-	2	-	-	-	-	2	-	15	-	-	-
ʒ	6	-	3	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	1	4	-	-	-
ʂ	5	-	1	-	5	-	1	-	-	-	-	3	-	2	-	-	-	1	-	-	-	-	19	1
ʐ	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	1	-

Table B-41 Kalasha: Observed (O) and Expected (E) values with O/E ratios for pairs of non-adjacent coronal obstruents co-occurring in *roots*; segments classified by place and manner; linear order of segments collapsed (n = 766)

C_1\\C_2		t, t^h, d, d^h	ʈ, ʈ^h, ɖ, ɖ^h	ʦ, ʦ^h, ʣ	ʧ, ʧ^h ʤ, ʤ^h	ʈʂ, ʈʂ^h, ɖʐ	s, z	ʃ, ʒ	ʂ, ʐ
t, t^h, d, d^h	O	70	7	12	67	22	143	61	24
	E	73.6	56.2	8.5	73.9	20.4	92.4	44.3	33.0
	O/E	0.95	0.12	1.41	0.91	1.08	1.55	1.38	0.73
ʈ, ʈ^h, ɖ, ɖ^h	O		39	4	42	1	22	21	7
	E		10.7	3.3	28.5	7.8	35.4	16.9	12.5
	O/E		3.64	1.22	1.47	0.13	0.62	1.24	0.56
ʦ, ʦ^h, ʣ	O			3	1	0	3	1	0
	E			0.2	3.9	1.2	5.2	2.5	2.0
	O/E			13.06	0.26	0.00	0.58	0.39	0.00
ʧ, ʧ^h ʤ, ʤ^h	O				16	1	48	13	29
	E				16.0	10.2	44.8	22.0	17.7
	O/E				1.00	0.10	1.07	0.59	1.64
ʈʂ, ʈʂ^h, ɖʐ	O					18	1	0	5
	E					1.4	12.8	6.1	4.6
	O/E					12.71	0.08	0.00	1.09
s, z	O						35	9	1
	E						28.8	27.7	21.1
	O/E						1.22	0.32	0.05
ʃ, ʒ	O							19	0
	E							6.7	10.0
	O/E							2.85	0.00
ʂ, ʐ	O								21
	E								3.6
	O/E								5.86

Table B-42 Kalasha: Observed (O) and Expected (E) values with O/E ratios for pairs of non-adjacent coronal obstruents co-occurring in *roots*; segments classified by place and sibilant/non-sibilant manner; linear order collapsed (n = 766)

$C_1 \backslash C_2$		t, t^h, d, d^h	ʈ, ʈ^h, ɖ, ɖ^h	ʦ, ʦ^h, ʥ, s, z	ʧ, ʧ^h, ʤ, ʤ^h, ʃ, ʒ	ʈʂ, ʈʂ^h, ɖʐ, ʂ, ʐ
t, t^h, d, d^h	O	70	7	155	128	46
	E	73.6	56.2	100.9	118.2	53.4
	O/E	0.95	0.12	1.54	1.08	0.86
ʈ, ʈ^h, ɖ, ɖ^h	O		39	26	63	8
	E		10.7	38.6	45.5	20.3
	O/E		3.64	0.67	1.39	0.39
ʦ, ʦ^h, ʥ, s, z	O			41	59	2
	E			34.2	79.0	37.0
	O/E			1.20	0.75	0.05
ʧ, ʧ^h, ʤ, ʤ^h, ʃ, ʒ	O				48	30
	E				44.6	44.0
	O/E				1.08	0.68
ʈʂ, ʈʂ^h, ɖʐ, ʂ, ʐ	O					44
	E					9.6
	O/E					4.59

Table B-43 Kalasha: Observed (O) and Expected (E) values with O/E ratios for pairs of non-adjacent coronal obstruents co-occurring in *roots*; segments classified as retroflex vs. non-retroflex and sibilant vs. non-sibilant; linear order of segments collapsed (n = 766)

$C_1 \backslash C_2$		t, t^h, d, d^h	ʈ, ʈ^h, ɖ, ɖ^h	ʦ, ʦ^h, ʥ, s, z ʧ, ʧ^h, ʤ, ʤ^h, ʃ, ʒ	ʈʂ, ʈʂ^h, ɖʐ, ʂ, ʐ
t, t^h, d, d^h	O	70	7	283	46
	E	73.6	56.2	219.2	53.4
	O/E	0.95	0.12	1.29	0.86
ʈ, ʈ^h, ɖ, ɖ^h	O		39	89	8
	E		10.7	84.1	20.3
	O/E		3.64	1.06	0.39
ʦ, ʦ^h, ʥ, s, z ʧ, ʧ^h, ʤ, ʤ^h, ʃ, ʒ	O			148	32
	E			157.8	81.1
	O/E			0.94	0.39
ʈʂ, ʈʂ^h, ɖʐ, ʂ, ʐ	O				44
	E				9.6
	O/E				4.59

Kharia

Classification: Munda, South

Descriptive Source(s): Biligiri (1965); Peterson (2008)

Consonant Phonemes:

p	t	ʈ	ʧ	k	(ʔ)
p^h	t^h	ʈ^h	ʧ^h	k^h	
b	d	ɖ	ʤ	ɡ	
b^h	d^h	ɖ^h	ʤ^h	ɡ^h	
	s				h
m	n	(ɳ)	ɲ	ŋ	
	r	(ɽ) (ɽ^h)			
	l				
w			j		

Lexical Data Source: Peterson (2009)

Search Domain: headwords containing #C_1V(N/ʔ)C_2 sequences

Segment Class: $\{C_1, C_2\}$ = coronal plosives and retroflex sonorants

Table B-44 Kharia: Observed values for each $C_1 \ldots C_2$ pair (n = 128)

C_1\ C_2	t	t^h	d	d^h	ʈ	ʈ^h	ɖ	ɖ^h	ɳ	ɽ	ɽ^h
t	7	1	-	-	-	-	9	-	-	6	-
t^h	-	2	-	-	-	-	-	-	-	1	-
d	5	-	11	2	-	-	-	-	1	4	1
d^h	3	-	-	4	1	-	-	-	-	2	-
ʈ	-	-	-	-	9	-	3	1	-	3	1
ʈ^h	-	-	-	-	1	7	1	2	1	2	1
ɖ	-	-	-	-	-	-	10	-	3	9	2
ɖ^h	-	-	-	-	1	3	-	5	-	2	1
ɳ	-	-	-	-	-	-	-	-	-	-	-
ɽ	-	-	-	-	-	-	-	-	-	-	-
ɽ^h	-	-	-	-	-	-	-	-	-	-	-

Table B-45 Kharia: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner (n = 128)

C_1\ C_2		t, t^h, d, d^h	ʈ, ʈ^h, ɖ, ɖ^h	ɳ, ɽ, ɽ^h
t, t^h, d, d^h	O	35	10	15
	E	16.4	24.8	18.8
	O/E	2.13	0.40	0.80
ʈ, ʈ^h, ɖ, ɖ^h	O	0	43	25
	E	18.6	28.2	21.3
	O/E	0.00	1.53	1.18

Konda

Classification: Dravidian, South-Central

Descriptive Source(s): Krishnamurti (1969); Krishnamurti & Benham (1998)

Consonant Phonemes:

p	t	ʈ		k	ʔ
b	d	ɖ		g	
	s				(h)
	z				
m	n	ɳ		ŋ	
	r, ɾ	ɽ			
	r̥				
	l				
v			j		

Lexical Data Source: Krishnamurti (1969)

Search Domain: headwords containing $\#C_1V(N)C_2$ sequences

Segment Class: $\{C_1, C_2\}$ = coronal plosives and retroflex sonorants

Table B-46 Konda: Observed values for each $C_1 \ldots C_2$ pair (n = 37)

$C_1 \backslash C_2$	t	d	ʈ	ɖ	ɳ	ɽ
t	1	-	-	-	-	8
d	1	3	-	-	-	3
ʈ	-	-	3	4	-	4
ɖ	-	-	2	7	1	-
ɳ	-	-	-	-	-	-
ɽ	-	-	-	-	-	-

Table B-47 Konda: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner (n = 37)

$C_1 \backslash C_2$		t, d	ʈ, ɖ	ɳ, ɽ
	O	5	0	11
t, d	E	2.2	6.9	6.9
	O/E	2.31	0.00	1.59
	O	0	16	5
ʈ, ɖ	E	2.8	9.1	9.1
	O/E	0.00	1.76	0.55

Konkani

Classification:	Indo-Aryan, Southern					
Descriptive Source(s):	Miranda (2003)					
Consonant Phonemes:[5]	p	t	ʈ		k	
	p^{h} [f]	t^{h}	ʈ^{h}		k^{h}	
	b	d	ɖ		g	
	b^{h}	d^{h}	ɖ^{h}		g^{h}	
		ʦ		ʧ		
		ʦ^{h}		ʧ^{h}		
		ʣ		ʤ		
		ʣ^{h}		ʤ^{h}		
		s		ʃ		h
	m	n	ɳ			
	m^{h}	n^{h}				
		r				
		l	ɭ			
		l^{h}				
	w			j		
	w^{h}			j^{h}		
Lexical Data Source:	Maffei (1883)					
Search Domain:	headwords containing #$C_1V(N)C_2$ sequences					
Segment Class:	$\{C_1, C_2\}$ = coronal plosives and retroflex sonorants					

Table B-48 Konkani: Observed values for each $C_1 \ldots C_2$ pair (n = 127)

$C_1 \backslash C_2$	t	t^{h}	d	d^{h}	ʈ	ʈ^{h}	ɖ	ɖ^{h}	ɳ	ɭ
t	12	-	4	-	8	-	12	-	2	19
t^{h}	-	-	-	-	1	-	3	-		1
d	15	-	9	-	4	-	18	-	8	6
d^{h}	1	-	-	-	-	-	2	-	-	1
ʈ	-	-	-	-	-	-	-	-	-	-
ʈ^{h}	-	-	-	-	-	-	-	-	1	-
ɖ	-	-	-	-	-	-	-	-	-	-
ɖ^{h}	-	-	-	-	-	-	-	-	-	-
ɳ	-	-	-	-	-	-	-	-	-	-
ɭ	-	-	-	-	-	-	-	-	-	-

[5] Secondary palatalization is also phonemic for all consonants except affricates, aspirated sonorants and /j/.

Table B-49 Konkani: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner (n = 127)

C_1\ C_2		t, t^h, d, d^h	ʈ, ʈ^h, ɖ, ɖ^h	ɳ, ɭ
t, t^h, d, d^h	O	41	48	37
	E	40.7	47.6	37.7
	O/E	1.01	1.01	0.98
ʈ, ʈ^h, ɖ, ɖ^h	O	0	0	1
	E	0.3	0.4	0.3
	O/E	0.00	0.00	3.34

Korwa

Classification: Munda, North

Descriptive Source(s): George & Joseph (2008)

Consonant Phonemes:

p	t	ʈ	tʃ	k	ʔ
(p^h)	(t^h)	(ʈ^h)	(tʃ^h)	(k^h)	
b	d	ɖ	ʤ	g	
(b^h)	(d^h)	(ɖ^h)	(ʤ^h)	(g^h)	
	s				h
m	n	(ɳ)		ŋ	
	r	ɽ			
	l				
w			j		

Lexical Data Source: Binzy Joseph (personal communication, Toolbox database)

Search Domain: headwords containing #C_1V(N)C_2 sequences

Segment Class: {C_1, C_2} = coronal plosives and retroflex sonorants

Table B-50 Korwa: Observed values for each $C_1 \ldots C_2$ pair (n = 49)

C_1\ C_2	t	t^h	d	d^h	ʈ	ʈ^h	ɖ	ɖ^h	ɽ
t	7	-	2	-	-	-	-	-	2
t^h	-	-	-	-	-	-	-	-	-
d	1	-	2	-	-	-	-	-	4
d^h	1	-	1	1	-	-	-	-	-
ʈ	-	-	-	-	5	-	4	-	3
ʈ^h	-	-	-	-	-	-	-	-	3
ɖ	-	-	-	-	-	1	3	2	3
ɖ^h	-	-	-	-	-	-	-	2	2
ɽ	-	-	-	-	-	-	-	-	-

Table B-51 Korwa: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner (n = 49)

C_1\ C_2		t, t^h, d, d^h	ʈ, ʈ^h, ɖ, ɖ^h	ɽ
t, t^h, d, d^h	O	15	0	6
	E	6.4	7.3	7.3
	O/E	2.33	0.00	0.82
ʈ, ʈ^h, ɖ, ɖ^h	O	0	17	11
	E	8.6	9.7	9.7
	O/E	0.00	1.75	1.13

Kui

Classification: Dravidian, South-Central

Descriptive Source(s): Winfield (1928); cf. Burrow & Bhattacharya (1961)

Consonant Phonemes:

p	t	ʈ		k	(ʔ)
b	d	ɖ	ʤ	g	
	s				h
m	n	ɳ			
	r	ɽ			
	l				
v					

Lexical Data Sources: Winfield (1928); Burrow & Bhattacharya (1961); Burrow & Emeneau (1984)[6]

Search Domain: headwords containing $\#C_1(r)V(N)C_2$ sequences

Segment Class: $\{C_1, C_2\}$ = coronal plosives and retroflex sonorants

Table B-52 Kui: Observed values for each $C_1 \ldots C_2$ pair (n = 45)

C_1\ C_2	t	d	ʈ	ɖ	ɳ	ɽ
t	1	1	2	1	-	7
d	1	-	-	-	-	-
ʈ	-	-	3	7	2	-
ɖ	-	-	2	13	-	2
ɳ	-	-	-	-	-	-
ɽ	-	2	-	1	-	-

6 No individual source contained enough data for an analysis of Kui, so data was combined from three sources. Where the same word was listed in more than one source it was counted only once.

Table B-53 Kui: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner (n = 42)[7]

$C_1 \backslash C_2$		t, d	ʈ, ɖ	ɳ, ɽ
t, d	O	3	3	7
	E	0.9	8.7	3.4
	O/E	3.23	0.35	2.06
ʈ, ɖ	O	0	25	4
	E	2.1	19.3	7.6
	O/E	0.00	1.29	0.53

Kumauni

Classification: Indo-Aryan, Northern

Descriptive Source(s): Apte & Pattanayak (1967); van Riezen (2000)

Consonant Phonemes:[8]

p	t	ʈ	tʃ	k	
p^h	t^h	$ʈ^h$	$tʃ^h$	k^h	
b	d	ɖ	dʒ	g	
b^h	d^h	$ɖ^h$	$dʒ^h$	g^h	
	s				h
m	n	ɳ		ŋ	
	r	[ɽ]			
	l				
w			j		

Lexical Data Source: Irene van Riezen (personal communication, lexical database)

Search Domain: headwords containing $\#C_1V(N)C_2$ sequences

Segment Class: $\{C_1, C_2\}$ = coronal plosives and retroflex sonorants

---

[7] Some of the South-Central Dravidian languages, including Kui, are unique among the languages of South Asia in allowing the retroflex flap /ɽ/ word-initially in a small number of words. In order to make the Kui data directly comparable to that of other languages, the few cases of initial /ɽ/ are excluded from Table B-53. The observed counts for /ɽ…d/ and /ɽ…ɖ/ suggest that initial /ɽ/ does not trigger retroflex consonant harmony in following coronal plosives but co-occurs with them in much the same way as non-initial /ɽ/.

[8] [ɽ] is regarded as an allophone of /ɖ/. Apte & Pattanayak (1967) treat aspirates as clusters of C + /h/.

Table B-54 Kumauni: Observed values for each $C_1 \ldots C_2$ pair ($n = 54$)

C_1\ C_2	t	t^h	d	d^h	ʈ	ʈ^h	ɖ	ɖ^h	ɳ	ɽ
t	6	1	-	-	-	-	1	-	5	2
t^h	-	-	-	-	-	-	-	-	-	2
d	3	-	6	1	-	-	1	-	2	1
d^h	3	-	-	2	-	-	-	-	2	-
ʈ	-	-	-	-	3	-	-	-	-	2
ʈ^h	-	-	-	-	-	-	2	-	-	1
ɖ	-	-	-	-	1	-	2	-	1	3
ɖ^h	-	-	-	-	1	-	-	-	-	-
ɳ	-	-	-	-	-	-	-	-	-	-
ɽ	-	-	-	-	-	-	-	-	-	-

Table B-55 Kumauni: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner ($n = 54$)

C_1\ C_2		t, t^h, d, d^h	ʈ, ʈ^h, ɖ, ɖ^h	ɳ, ɽ
t, t^h, d, d^h	O	22	2	14
	E	15.5	7.7	14.8
	O/E	1.42	0.26	0.95
ʈ, ʈ^h, ɖ, ɖ^h	O	0	9	7
	E	6.5	3.3	6.2
	O/E	0.00	2.76	1.13

Kurux

Classification: Dravidian, Northern

Descriptive Source(s): Pfeiffer (1972); cf. Hanh (1911); Krishnamurti (2003, pp. 74–75)

Consonant Phonemes:

p	t	ʈ	ʧ	k	ʔ
p^h	t^h	ʈ^h	ʧ^h	k^h	
b	d	ɖ	ʤ	g	
b^h	d^h	ɖ^h	ʤ^h	g^h	
	s			x	h
m	n	(ɳ)		ŋ	
	r	ɽ			
		ɽ^h			
	l				
w			j		

Lexical Data Source:	Grignard ([1924] 1986)
Search Domain:	headwords containing #$C_1V(N)C_2$ sequences
Segment Class:	$\{C_1, C_2\}$ = coronal plosives and retroflex sonorants

Table B-56 Kurux: Observed values for each $C_1 \ldots C_2$ pair (n = 286)

$C_1 \backslash C_2$	t	t^h	d	d^h	ʈ	ʈ^h	ɖ	ɖ^h	ɽ	ɽ^h
t	23	-	8	-	-	-	1	-	24	2
t^h	1	12	-	-	-	-	-	-	2	-
d	7	-	18	6	-	-	1	-	12	5
d^h	6	-	1	9	-	-	-	-	16	1
ʈ	-	-	1	-	25	3	6	-	16	1
ʈ^h	-	-	-	-	2	14	-	-	2	-
ɖ	-	-	-	-	5	1	22	4	6	4
ɖ^h	-	-	-	-	3	4	1	6	-	5
ɽ	-	-	-	-	-	-	-	-	-	-
ɽ^h	-	-	-	-	-	-	-	-	-	-

Table B-57 Kurux: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner (n = 286)

$C_1 \backslash C_2$		t, t^h, d, d^h	ʈ, ʈ^h, ɖ, ɖ^h	ɽ, ɽ^h
t, t^h, d, d^h	O	91	2	62
	E	49.9	53.1	52.0
	O/E	1.83	0.04	1.19
ʈ, ʈ^h, ɖ, ɖ^h	O	1	96	34
	E	42.1	44.9	44.0
	O/E	0.02	2.14	0.77

Kuvi

Classification:	Dravidian, South-Central
Descriptive Source(s):	Israel (1979); cf. Reddy, Upadhyaya, & Reddy (1974)

Consonant Phonemes:

p	t	ʈ	ʧ	k	ʔ
b	d	ɖ	ʤ	ɡ	
	s				h
m	n	ɳ		ŋ	
	r	ɽ			
	l				
v			j		

Lexical Data Source: Israel (1979)

Search Domain: headwords containing $\#C_1V(N)C_2$ sequences

Segment Class: $\{C_1, C_2\}$ = coronal plosives and retroflex sonorants

Table B-58 Kuvi: Observed values for each $C_1 \ldots C_2$ pair (n = 83)

$C_1 \backslash C_2$	t	d	ʈ	ɖ	ɳ	ɽ
t	4	4	-	4	4	6
d	1	2	-	1	1	5
ʈ	-	-	11	13	4	1
ɖ	-	-	2	15	-	5
ɳ	4	4	-	4	4	6
ɽ	1	2	-	1	1	5

Table B-59 Kuvi: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner (n = 83)

$C_1 \backslash C_2$		t, d	ʈ, ɖ	ɳ, ɽ
t, d	O	11	5	16
	E	4.2	17.7	10.0
	O/E	2.59	0.28	1.60
ʈ, ɖ	O	0	41	10
	E	6.8	28.3	16.0
	O/E	0.00	1.45	0.63

Malto

Classification: Dravidian, Northern

Descriptive Source(s): Mahapatra (1979); cf. Das (1973); Steever (1998d)

Consonant Phonemes:

p	t		ʈ	ʧ	k	q	
b	d		ɖ	ʤ	g		
	ð	s				ʁ	h
m	n			ɲ	ŋ		
		l					
		r	ɽ				
w				j			

Lexical Data Source: Mahapatra (1987)

Search Domain 1: headwords containing $\#C_1V(N)C_2$ sequences

Segment Class 1: $\{C_1, C_2\}$ = coronal plosives and retroflex sonorants

Search Domain 2: headwords containing $\#C_1 \ldots C_2$ sequences

Segment Class 2: $\{C_1, C_2\}$ = dorsal consonants

Table B-60 Malto: Observed values for co-occurring coronal plosives and retroflex sonorants in #$C_1V(N)C_2$ sequences (n = 137)

C_1\ C_2	t	d	ʈ	ɖ	ɽ
t	14	4	-	-	20
d	2	10	-	-	17
ʈ	-	-	22	15	5
ɖ	-	-	4	19	5
ɽ	-	-	-	-	-

Table B-61 Malto: Observed (O) and Expected (E) values with O/E ratios for co-occurring coronal plosives and retroflex sonorants; segments classified by place and manner (n = 137)

C_1\ C_2		t, d	ʈ, ɖ	ɽ
t, d	O	30	0	37
	E	14.7	29.3	23.0
	O/E	2.04	0.00	1.61
ʈ, ɖ	O	0	60	10
	E	15.3	30.7	24.0
	O/E	0.00	1.96	0.42

Table B-62 Malto: Observed values for co-occurring dorsal consonants in #C_1...C_2 sequences (n = 132)

C_1\ C_2	k	g	q	ʁ	ŋ
k	44	25	-	-	3
g	1	13	-	-	10
q	-	-	17	19	-
ʁ	-	-	-	-	-
ŋ	-	-	-	-	-

Table B-63 Malto: Observed (O) and Expected (E) values with O/E ratios for co-occurring dorsal consonants; segments classified by place and manner (n = 132)

C_1\ C_2		k, g	q, ʁ	ŋ
k, g	O	83	0	13
	E	60.4	26.2	9.5
	O/E	1.38	0.00	1.38
q, ʁ	O	0	36	0
	E	22.6	9.8	3.5
	O/E	0.00	3.67	0.00

Marathi

Classification: Indo-Aryan, Southern

Descriptive Source(s): Pandharipande (1997; 2003); Wali (2005)

Consonant Phonemes:

p	t	ʈ		k	
pʰ	tʰ	ʈʰ		kʰ	
b	d	ɖ		g	
bʰ	dʰ	ɖʰ		gʰ	
	ʦ		ʧ		
			ʧʰ		
	ʣ		ʤ		
	ʣʰ		ʤʰ		
	s		ʃ		h
m	n	ɳ			
	r				
	l	ɭ			
w			j		

Lexical Data Source: Molesworth (1857)

Search Domain: headwords containing $\#C_1V(N)C_2$ sequences

Segment Class: $\{C_1, C_2\}$ = coronal plosives and retroflex sonorants

Table B-64 Marathi: Observed values for each $C_1 \ldots C_2$ pair (n = 1833)

$C_1 \backslash C_2$	t	tʰ	d	dʰ	ʈ	ʈʰ	ɖ	ɖʰ	ɳ	ɭ
t	93	25	27	6	91	13	221	5	42	112
tʰ	9	4	1	-	30	-	43	-	2	22
d	110	3	37	41	21	8	154	30	51	44
dʰ	62	-	34	11	22	-	131	-	27	16
ʈ	-	-	-	-	14	-	-	-	38	45
ʈʰ	-	-	-	-	3	11	2	-	34	2
ɖ	1	1	-	-	-	-	7	-	4	61
ɖʰ	-	-	-	-	-	-	5	13	18	26
ɳ	-	-	-	-	-	-	-	-	-	-
ɭ	-	-	-	-	-	-	-	-	-	-

Table B-65 Marathi: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner (n = 1833)

C_1\ C_2		t, t^h, d, d^h	ʈ, ʈ^h, ɖ, ɖ^h	ɳ, ɭ
t, t^h, d, d^h	O	463	769	316
	E	392.7	695.9	459.4
	O/E	1.18	1.11	0.69
ʈ, ʈ^h, ɖ, ɖ^h	O	2	55	228
	E	72.3	128.1	84.6
	O/E	0.03	0.43	2.70

Mundari

Classification: Munda, North

Descriptive Source(s): Osada (2008); cf. Cook (1965); Gumperz (1957)

Consonant Phonemes:[9]

p	t	ʈ	ʧ	k	ʔ
(p^h)	(t^h)	(ʈ^h)	(ʧ^h)	(k^h)	
b	d	ɖ	ʤ	g	
(b^h)	(d^h)	(ɖ^h)	(ʤ^h)	(g^h)	
	s				h
m	n	(ɳ)	ñ	ŋ	
	r	ɽ			
		(ɽ^h)			
	l				
w			j		

Lexical Data Source: Bhaduri (1983 [1931])

Search Domain: headwords containing #C_1V(N)C_2 sequences

Segment Class: {C_1, C_2} = coronal plosives and retroflex sonorants

---

[9] Aspirated consonants are not listed as phonemes in most accounts but are included in Bhaduri's (1983 [1931]) data. They probably reflect Indo-Aryan loanwords. Osada (2008) identifies /ɳ/ as a phoneme, but it is not distinguished in Bhaduri's data.

Table B-66 Mundari: Observed values for each $C_1 \ldots C_2$ pair (n = 198)

$C_1 \backslash C_2$	t	tʰ	d	dʰ	ʈ	ʈʰ	ɖ	ɖʰ	ɽ	ɽʰ
t	6	1	6	-	2	-	-	-	13	-
tʰ	1	1	-	-	-	-	-	-	2	-
d	6	1	18	3	2	-	1	-	20	-
dʰ	4	-	7	2	-	-	1	-	4	-
ʈ	1	-	1	-	16	1	9	-	3	1
ʈʰ	-	-	-	-	3	1	-	-	-	-
ɖ	-	-	2	-	7	-	26	-	7	3
ɖʰ	-	1	-	-	3	-	5	1	4	2
ɽ	-	-	-	-	-	-	-	-	-	-
ɽʰ	-	-	-	-	-	-	-	-	-	-

Table B-67 Mundari: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner (n = 198)

$C_1 \backslash C_2$		t, tʰ, d, dʰ	ʈ, ʈʰ, ɖ, ɖʰ	ɽ, ɽʰ
t, tʰ, d, dʰ	O	56	6	39
	E	31.1	39.8	30.1
	O/E	1.80	0.15	1.30
ʈ, ʈʰ, ɖ, ɖʰ	O	5	72	20
	E	29.9	38.2	28.9
	O/E	0.17	1.88	0.69

Nepali

Classification:	Indo-Aryan, Northern					
Descriptive Source(s):	Acharya (1991); Riccardi (2003); Khatiwada (2009)					
Consonant Phonemes:[10]	p	t	ʈ	ʧ	k	
	pʰ	tʰ	ʈʰ	ʧʰ	kʰ	
	b	d	ɖ	ʤ	ɡ	
	bʰ	dʰ	ɖʰ	ʤʰ	ɡʰ	
		s				h
	m	n	[ɳ]		ŋ	
		r	[ɽ] [ɽʰ]			
		l				
	w			j		

[10] [ɳ, ɽ, ɽʰ] are not phonemic but are distinguished orthographically in Turner (1931).

Lexical Data Source: Turner (1931)
Search Domain: headwords containing #$C_1V(N)C_2$ sequences
Segment Class: $\{C_1, C_2\}$ = coronal plosives and retroflex sonorants

Table B-68 Nepali: Observed values for each $C_1 \ldots C_2$ pair (n = 597)

$C_1 \backslash C_2$	t	t^h	d	d^h	ʈ	ʈ^h	ɖ	ɖ^h	ɳ	ɽ	ɽ^h
t	67	5	20	3	4	1	0	-	-	19	-
t^h	19	-	1	-	-	-	-	-	-	3	-
d	30	-	30	13	3	2	10	2	1	13	4
d^h	30	-	5	6	1	-	-	-	-	6	-
ʈ	-	-	-	-	44	13	2	-	3	13	5
ʈ^h	-	-	-	-	43	-	2	-	1	21	-
ɖ	-	-	-	-	4	15	6	-	12	38	17
ɖ^h	-	-	-	-	22	1	5	-	4	28	-
ɳ	-	-	-	-	-	-	-	-	-	-	-
ɽ	-	-	-	-	-	-	-	-	-	-	-
ɽ^h	-	-	-	-	-	-	-	-	-	-	-

Table B-69 Nepali: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner (n = 597)

$C_1 \backslash C_2$		t, t^h, d, d^h	ʈ, ʈ^h, ɖ, ɖ^h	ɳ, ɽ, ɽ^h
t, t^h, d, d^h	O	229	23	46
	E	114.3	89.8	93.8
	O/E	2.00	0.26	0.49
ʈ, ʈ^h, ɖ, ɖ^h	O	0	157	142
	E	114.7	90.2	94.2
	O/E	0.00	1.74	1.51

Oriya

Classification: Indo-Aryan, Eastern

Descriptive Source(s): Ray (2003)

Consonant Phonemes:[11]

p	t	ʈ	ʧ	k	
p^h	t^h	ʈ^h	ʧ^h	k^h	
b	d	ɖ	ʤ	g	
b^h	d^h	ɖ^h	ʤ^h	g^h	
	s				h
m	n	ɳ		(ŋ)	
	r	[ɽ]			
		[ɽ^h]			
	l	ɭ			
w			j		

Lexical Data Source: Turner (1969)

Search Domain: words containing $\#C_1V(N)C_2$ sequences[12]

Segment Class: $\{C_1, C_2\}$ = coronal plosives and retroflex sonorants

Table B-70 Oriya: Observed values for each $C_1 \ldots C_2$ pair ($n = 145$)

$C_1 \backslash C_2$	t	t^h	d	d^h	ʈ	ʈ^h	ɖ	ɖ^h	ɳ	ɽ	ɽ^h	ɭ
t	11	1	-	-	1	1	1	-	1	10	-	11
t^h	3	-	-	-	1	-	1	-	1	2	-	4
d	5	-	7	2	-	2	3	-	5	6	4	8
d^h	1	-	1	-	1	-	3	-	1	7	-	3
ʈ	-	-	-	-	5	-	1	-	3	1	-	3
ʈ^h	-	-	-	-	3	-	1	-	3	1	-	-
ɖ	-	-	-	-	1	-	1	-	3	1	-	3
ɖ^h	-	-	-	-	-	-	1	-	2	-	-	4
ɳ	-	-	-	-	-	-	-	-	-	-	-	-
ɽ	-	-	-	-	-	-	-	-	-	-	-	-
ɽ^h	-	-	-	-	-	-	-	-	-	-	-	-
ɭ	-	-	-	-	-	-	-	-	-	-	-	-

[11] [ɽ, ɽ^h] are distinguished in Turner's (1969) transcription, but regarded as allophones of /ɖ, ɖ^h/ in Ray (2003).

[12] One word form was counted per etymological group in Turner (1969). In the event that both harmonic and disharmonic forms are included in the same etymological group, the harmonic form was counted.

Table B-71 Oriya: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner (n = 145)

C_1\ C_2		t, t^h, d, d^h	ʈ, ʈ^h, ɖ, ɖ^h	ɳ, ɽ, ɽ^h, ɭ
t, t^h, d, d^h	O	31	14	63
	E	23.1	20.1	64.8
	O/E	1.34	0.70	0.97
ʈ, ʈ^h, ɖ, ɖ^h	O	0	13	24
	E	7.9	6.9	22.2
	O/E	0.00	1.89	1.08

Pāli

Classification: Middle Indo-Aryan (600 BCE – 1000 CE)

Descriptive Source(s): Oberlies (2003)

Consonant Phonemes:[13]

p	t	ʈ	ʧ	k	
p^h	t^h	ʈ^h	ʧ^h	k^h	
b	d	ɖ	ʤ	ɡ	
b^h	d^h	ɖ^h	ʤ^h	ɡ^h	
	s				h
m	n	ɳ	ɲ	ŋ	
	r				
	l	ɭ			
		ɭ^h			
v			j		

Lexical Data Source: Pali Text Society (1921–1925)

Search Domain: headwords containing #C_1V(N)C_2 sequences

Segment Class: {C_1, C_2} = coronal plosives and retroflex sonorants

[13] Some instances of /ɳ/ are followed by /h/ in the data source (Pali Text Society, 1921–1925). These might be interpreted as /ɳ^h/, though /ɳ^h/ is not identified as a phoneme in Oberlies (2003). Counts for /ɳh/ and /ɳ/ are both included under /ɳ/ here.

Table B-72 Pāli: Observed values for each $C_1 \ldots C_2$ pair (n = 202)

C_1\ C_2	t	t^h	d	d^h	ʈ	ʈ^h	ɖ	ɖ^h	ɳ	ɭ	ɭ^h
t	30	11	14	1	4	-	7	-	14	5	-
t^h	3	-	-	1	-	-	1	-	3	-	-
d	20	-	15	9	-	17	6	2	6	4	2
d^h	19	-	-	-	-	-	-	-	-	-	-
ʈ	-	-	-	-	-	-	-	-	-	-	-
ʈ^h	5	-	-	-	1	-	-	-	-	-	-
ɖ	1	-	-	-	-	1	-	-	-	-	-
ɖ^h	-	-	-	-	-	-	-	-	-	-	-
ɳ	-	-	-	-	-	-	-	-	-	-	-
ɭ	-	-	-	-	-	-	-	-	-	-	-
ɭ^h	-	-	-	-	-	-	-	-	-	-	-

Table B-73 Pāli: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner (n = 202)

C_1\ C_2		t, t^h, d, d^h	ʈ, ʈ^h, ɖ, ɖ^h	ɳ, ɭ, ɭ^h
	O	123	37	34
t, t^h, d, d^h	E	123.9	37.5	32.7
	O/E	0.99	0.99	1.04
	O	6	2	0
ʈ, ʈ^h, ɖ, ɖ^h	E	5.1	1.5	1.3
	O/E	1.17	1.29	0.00

Palula

Classification: Indo-Aryan, Northwest, Dardic

Descriptive Source(s): Liljegren (2008); Liljegren & Haider (2009)

Consonant Phonemes:[14]	p	t	ʈ		k	(q)	
	(p^{h})	(t^{h})	(ʈ^{h})		(k^{h})		
	b	d	ɖ		ɡ		
	(b^{h})	(d^{h})	(ɖ^{h})		(ɡ^{h})		
		ʦ	ʈʂ	ʧ			
		(ʦ^{h})	(ʈʂ^{h})	(ʧ^{h})			
	(f)	s	ʂ	ʃ	x		h
		z	(ʐ)	ʒ	ɣ		
		(z^{h})	(ʐ^{h})	(ʒ^{h})			
	m	n	ɳ				
	(m^{h})	(n^{h})					
		r	ɽ				
		(r^{h})					
		l					
		(l^{h})					
	w			j			
	(w^{h})			(j^{h})			

Lexical Data Source: Henrik Liljegren (personal communication, Toolbox database)

Search Domain: headwords containing #C_1V(N)C_2 sequences

Segment Class: $\{C_1, C_2\}$ = coronal obstruents (plosives, affricates, fricatives)

[14] Liljegren (2008) and Liljegren & Haider (2009) treat aspiration as a suprasegmental feature applying to lexical stems as opposed to segments. There is never more than one aspirate in a word, and it is typically word-initial.

Table B-74 Palula: Observed values for each $C_1 \ldots C_2$ pair (n = 113)

C_1\\C_2	t	tʰ	d	dʰ	ʈ	ʈʰ	ɖ	ɖʰ	ts	tsʰ	tʃ	tʃʰ	ʈʂ	ʈʂʰ	s	z	zʰ	ʃ	ʒ	ʒʰ	ʂ	ʐ	ʐʰ
t	1	-	2	-	-	-	-	-	-	-	-	-	1	-	6	1	-	1	1	-	-	-	-
tʰ	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-
d	1	-	5	-	-	-	-	-	-	-	-	-	-	1	2	-	-	6	1	-	2	-	-
dʰ	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ʈ	-	-	-	-	3	-	1	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-
ʈʰ	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ɖ	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-
ɖʰ	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ts	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
tsʰ	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
tʃ	1	-	2	-	2	-	1	-	-	-	2	-	-	-	-	1	-	-	-	-	1	-	-
tʃʰ	1	-	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-
ʈʂ	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
ʈʂʰ	2	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-
s	11	1	1	-	-	1	2	-	-	-	1	-	-	-	3	-	-	-	-	-	-	-	-
z	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
zʰ	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ʃ	1	-	3	-	1	-	1	-	-	-	-	-	-	-	-	-	-	4	2	-	-	-	-
ʒ	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	1	-	-
ʒʰ	-	-	1	-	4	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	1	-	-
ʂ	-	-	1	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	1	-	-
ʐ	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
ʐʰ	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-

Table B-75 Palula: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner; linear order of segments collapsed (n = 113)

C_1\ C_2		t, t^h, d, d^h	ʈ, ʈ^h, ɖ, ɖ^h	ʦ, ʦ^h	ʧ, ʧ^h	tʂ, tʂ^h	s, z, z^h	ʃ, ʒ, ʒ^h	ʂ, ʐ, ʐ^h
t, t^h, d, d^h	O	11	0	0	5	4	25	16	3
	E	12.3	10.2	0.4	6.6	2.4	12.8	13.5	4.5
	O/E	0.89	0.00	0.00	0.76	1.70	1.95	1.19	0.66
ʈ, ʈ^h, ɖ, ɖ^h	O		8	1	5	0	4	6	0
	E		1.9	0.2	3.2	1.0	5.7	6.0	1.8
	O/E		4.11	5.14	1.58	0.00	0.70	1.00	0.00
ʦ, ʦ^h	O			0	0	0	0	0	0
	E			0.0	0.0	0.0	0.1	0.2	0.1
	O/E			0.00	0.00	0.00	0.00	0.00	0.00
ʧ, ʧ^h	O				2	0	2	1	2
	E				0.6	0.5	3.0	3.1	1.3
	O/E				3.23	0.00	0.68	0.32	1.50
tʂ, tʂ^h	O					1	0	0	1
	E					0.1	1.2	1.2	0.5
	O/E					9.42	0.00	0.00	2.22
s, z, z^h	O						3	1	0
	E						3.1	6.6	2.5
	O/E						0.96	0.15	0.00
ʃ, ʒ, ʒ^h	O							7	2
	E							3.5	2.6
	O/E							2.02	0.77
ʂ, ʐ, ʐ^h	O								3
	E								0.4
	O/E								7.53

Table B-76 Palula: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and sibilant/non-sibilant manner; linear order collapsed (n = 113)

C$_1$\ C$_2$		t, t^h, d, d^h	ʈ, ʈ^h, ɖ, ɖ^h	ʦ, ʦ^h, s, z, z^h	ʧ, ʧ^h, ʃ, ʒ, ʒ^h	ʈʂ, ʈʂ^h, ʂ, ʐ, ʐ^h
t, t^h, d, d^h	O	11	0	25	21	7
	E	12.3	10.2	13.2	20.0	6.9
	O/E	0.89	0.00	1.90	1.05	1.02
ʈ, ʈ^h, ɖ, ɖ^h	O		8	5	11	0
	E		1.9	5.9	9.2	2.8
	O/E		4.11	0.85	1.20	0.00
ʦ, ʦ^h, s, z, z^h	O			3	3	0
	E			3.3	9.7	3.7
	O/E			0.92	0.31	0.00
ʧ, ʧ^h, ʃ, ʒ, ʒ^h	O				10	4
	E				7.2	5.7
	O/E				1.39	0.70
ʈʂ, ʈʂ^h, ʂ, ʐ, ʐ^h	O					5
	E					1.0
	O/E					5.23

Table B-77 Palula: Observed (O) and Expected (E) values with O/E ratios; segments classified as retroflex/non-retroflex and sibilant/non-sibilant; linear order of segments collapsed (n = 113)

C$_1$\ C$_2$		t, t^h, d, d^h	ʈ, ʈ^h, ɖ, ɖ^h	ʦ, ʦ^h, s, z, z^h ʧ, ʧ^h, ʃ, ʒ, ʒ^h	ʈʂ, ʈʂ^h, ʂ, ʐ, ʐ^h
t, t^h, d, d^h	O	11	0	46	7
	E	12.3	10.2	33.2	6.9
	O/E	0.89	0.00	1.39	1.02
ʈ, ʈ^h, ɖ, ɖ^h	O		8	16	0
	E		1.9	15.0	2.8
	O/E		4.11	1.06	0.00
ʦ, ʦ^h, s, z, z^h ʧ, ʧ^h, ʃ, ʒ, ʒ^h	O			16	4
	E			20.2	9.4
	O/E			0.79	0.43
ʈʂ, ʈʂ^h, ʂ, ʐ, ʐ^h	O				5
	E				1.0
	O/E				5.23

Panjabi, Eastern

Classification:	Indo-Aryan, Central
Descriptive Source(s):	Bhatia (1993); Malik (1995); Shackle (2003); cf. Jain (1934)

Consonant Phonemes:[15]

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})	
(f)	s		ʃ	(x)	h
	(z)			(ɣ)	
m	n	ɳ			
	l	(ɭ)			
	r	ɽ			
		(ɽ^{h})			
w			j		

Lexical Data Source:	Goswami (2000)
Search Domain:	headwords containing $\#C_1V(N)C_2$ sequences
Segment Class:	$\{C_1, C_2\}$ = coronal plosives and retroflex sonorants

Table B-78 Panjabi: Observed values for each $C_1 \ldots C_2$ pair (n = 233)

C_1\ C_2	t	t^{h}	d	d^{h}	ʈ	ʈ^{h}	ɖ	ɖ^{h}	ɳ	ɽ	ɽ^{h}
t	15	2	8	-	3	-	1	-	6	26	-
t^{h}	1	5	3	-	-	-	-	-	2	7	-
d	7	-	13	4	-	-	1	-	3	8	3
d^{h}	6	-	6	-	-	-	-	-	3	12	-
ʈ	-	-	-	-	17	-	7	1	3	-	-
ʈ^{h}	-	-	-	-	-	18	4	-	4	-	-
ɖ	-	-	-	-	4	1	10	3	1	1	-
ɖ^{h}	-	-	-	-	1	2	10	-	1	-	-
ɳ	-	-	-	-	-	-	-	-	-	-	-
ɽ	-	-	-	-	-	-	-	-	-	-	-
ɽ^{h}	-	-	-	-	-	-	-	-	-	-	-

[15] Historic/orthographic voiced aspirates are pronounced unaspirated with accompanying pitch contours on neighbouring vowels. /ɭ/ is not phonemic in all dialects and is not distinguished orthographically.

Table B-79 Panjabi: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner (n = 233)

$C_1 \backslash C_2$		t, tʰ, d, dʰ	ʈ, ʈʰ, ɖ, ɖʰ	ɳ, ɽ, ɽʰ
t, tʰ, d, dʰ	O	70	5	70
	E	43.6	51.7	49.8
	O/E	1.61	0.10	1.41
ʈ, ʈʰ, ɖ, ɖʰ	O	0	78	10
	E	26.4	31.3	30.2
	O/E	0.00	2.49	0.33

Parji (Duruwa)

Classification: Dravidian, Central

Descriptive Source(s): Burrow & Bhattacharya (1953)

Consonant Phonemes:

p	t	ʈ	ʧ	k	
b	d	ɖ	ʤ	g	
	(s)				(h)
m	n		ɲ	ŋ	
	r	ɽ			
	l				
v			j		

Lexical Data Source: Burrow & Bhattacharya (1953)

Search Domain: headwords containing $\#C_1V(N)C_2$ sequences

Segment Class: $\{C_1, C_2\}$ = coronal plosives and retroflex sonorants

Table B-80 Parji: Observed values for each $C_1 \ldots C_2$ pair (n = 86)

$C_1 \backslash C_2$	t	d	ʈ	ɖ	ɽ
t	10	14	1	1	9
d	2	5	-	-	3
ʈ	-	-	8	13	-
ɖ	-	-	2	17	1
ɽ	-	-	-	-	-

Table B-81 Parji: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner (n = 86)

$C_1 \backslash C_2$		t, d	ʈ, ɖ	ɽ
t, d	O	31	2	12
	E	16.2	22.0	6.8
	O/E	1.91	0.09	1.76
ʈ, ɖ	O	0	40	1
	E	14.8	20.0	6.2
	O/E	0.00	2.00	0.16

Pengo

Classification: Dravidian, South-Central

Descriptive Source(s): Burrow & Bhattacharya (1970)

Consonant Phonemes:

p	t	ʈ	ʧ	k	
b	d	ɖ	ʤ	g	
	s				
	(z)				h
m	n	ɳ		ŋ	
	l				
	r	ɽ			
w			j		

Lexical Data Source: Burrow & Bhattacharya (1970)

Search Domain: headwords containing $\#C_1V(N)C_2$ sequences

Segment Class: $\{C_1, C_2\}$ = coronal plosives and retroflex sonorants

Table B-82 Pengo: Observed values for each $C_1 \ldots C_2$ pair (n = 77)

$C_1 \backslash C_2$	t	d	ʈ	ɖ	ɳ	ɽ
t	4	2	1	-	2	6
d	3	6	-	-	2	4
ʈ	-	-	7	11	2	7
ɖ	-	-	2	11	-	7
ɳ	-	-	-	-	-	-
ɽ	-	-	-	-	-	-

Table B-83 Pengo: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner (n = 77)

C_1\ C_2		t, d	ʈ, ɖ	ɳ, ɽ
t, d	O	15	1	14
	E	5.8	12.5	11.7
	O/E	2.57	0.08	1.20
ʈ, ɖ	O	0	31	16
	E	9.2	19.5	18.3
	O/E	0.00	1.59	0.87

Prakrit

Classification: Middle Indo-Aryan (600 BCE – 1000 CE)

Descriptive Source(s): Turner (1969); cf. Bubenik, 2003

Consonant Phonemes:[16]

p	t	ʈ	ʧ	k	
pʰ	tʰ	ʈʰ	ʧʰ	kʰ	
b	d	ɖ	ʤ	g	
bʰ	dʰ	ɖʰ	ʤʰ	gʰ	
	s				h
m		ɳ			
		ɳʰ			
	r				
	l				
	lʰ				
v			j		

[16] Prakrit (Pk.) is a cover term for a collection of MIA vernacular dialects. As a result, it is difficult to identify a single coherent phoneme inventory. The phonemes listed here are those found in Turner's (1969) word list, which includes data from the following MIA dialects: Śaurasenī, Paiśācī, Ardhamāgadhī, Māgadhī, Jaina Māgadhī, Mahārāṣṭrī and Jaina Mahārāṣṭrī (1969, p. vii). In addition to the phonemes identified here, Turner's data includes a few instances of /ʃ/ in the Māgadhī dialect (e.g., Māgadhī /maʃʧaliː/ 'fish' beside Pk. /maʧʧʰa/, CDIAL 9758; Māgadhī /ʃeː/ 'he, that' beside Ardhamāgadhī /seː/, CDIAL 12815), and at least one instance of /ʂ/ in a "Sanskritized" word form (i.e., /adʰjuʂʈa-/ 'three and a half' beside Pk. /aɖɖʰuʈʈʰa-/, CDIAL 649). All OIA coronal nasals were neutralized to orthographic ⟨ɳ⟩ in Prakrit, which may have represented phonetic [n] in word-initial position and retroflex [ɳ] elsewhere (Schwarzschild, 1973; Masica, 1991, p. 182). /n/ occurs in only a few words in Turner's data (e.g., Pk. /gilaːni-/ 'weariness', CDIAL 4401, beside Pk. /gilaːɳa-/ 'tired', CDIAL 4400; Pk. /ʤaːnaːveːi/, causative of /ʤaːɳaï/ 'knows', CDIAL 5193). Turner's data also includes a few instance of /ɭ/, usually alongside variants with /l/ (e.g., Pk. /soːɭasa, soːlasa/ 'sixteen', CDIAL 12812; Pk. /maɭaï/ 'rubs', CDIAL 10290).

Lexical Data Source:	Turner (1969)
Search Domain:	headwords containing $\#C_1V(N)C_2$ sequences
Segment Class:	$\{C_1, C_2\}$ = coronal plosives and retroflex sonorants

Table B-84 Prakrit: Observed values for each $C_1 \ldots C_2$ pair (n = 337)

$C_1 \backslash C_2$	t	t^h	d	d^h	ʈ	ʈ^h	ɖ	ɖ^h	ɳ	ɳ^h
t	26	4	11	1	2	4	28	-	26	3
t^h	1	-	-	1	1	-	3	1	13	-
d	16	-	8	5	-	9	11	4	13	-
d^h	7	1	2	1	-	2	5	-	14	-
ʈ	-	-	-	-	5	-	-	-	-	-
ʈ^h	-	-	-	-	-	2	1	1	3	-
ɖ	-	-	-	-	-	1	4	2	4	-
ɖ^h	-	-	-	-	-	-	-	5	-	-
ɳ	8	9	23	4	9	4	14	-	13	1
ɳ^h	-	-	-	-	-	-	-	-	1	-

Table B-85 Prakrit: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner (n = 337)

$C_1 \backslash C_2$		t, t^h, d, d^h	ʈ, ʈ^h, ɖ, ɖ^h	ɳ, ɳ^h
t, t^h, d, d^h	O	84	70	69
	E	84.7	78.1	60.2
	O/E	0.99	0.90	1.15
ʈ, ʈ^h, ɖ, ɖ^h	O	0	21	7
	E	10.6	9.8	7.6
	O/E	0.00	2.14	0.93
ɳ, ɳ^h (= [n, n^h]?)[17]	O	44	27	15
	E	32.7	30.1	23.2
	O/E	1.35	0.90	0.65

[17] Word-initial orthographic ⟨ɳ, ɳ^h⟩ may have represented phonetic [n, n^h] (Schwarzschild, 1973; Masica, 1991, p. 182).

Remo (Bonda)

Classification:	Munda, South					
Descriptive Source(s):	Alexander & Hannah (2000); Anderson & Harrison (2008a); Fernandez (1968)					
Consonant Phonemes:[18]	p	t	ʈ	(ʧ)	k	ʔ
	b	d	ɖ	ʤ	g	
		s				
		(z)				
	m	n	(ɳ)	ɲ	ŋ	
		r	(ɽ)			
		l				
	w			j		
Lexical Data Source:	Donegan & Stampe (2004)					
Search Domain:	headwords containing $\#C_1V(N)C_2$ sequences					
Segment Class:	$\{C_1, C_2\}$ = coronal plosives and retroflex sonorants					

Table B-86 Remo: Observed values for each $C_1 \ldots C_2$ pair (n = 53)

$C_1 \backslash C_2$	t	d	ʈ	ɖ	ɽ
t	9	1	-	3	4
d	4	4	-	2	1
ʈ	-	-	5	1	2
ɖ	1	-	1	14	1
ɽ	-	-	-	-	-

Table B-87 Remo: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner (n = 53)

$C_1 \backslash C_2$		t, d	ʈ, ɖ	ɽ
t, d	O	18	5	5
	E	10.0	13.7	4.2
	O/E	1.79	0.36	1.18
ʈ, ɖ	O	1	21	3
	E	9.0	12.3	3.8
	O/E	0.11	1.71	0.80

[18] The plains dialect may include /ʧ/ (Anderson & Harrison, 2008a). /z/ and /ɳ/ are included as phonemes in Fernandez (1964). /ɽ/ is phonemic in Alexander & Hannah (2000), but allophonic in other accounts. It is the only retroflex sonorant found in the search of the lexical data source.

Sanskrit

Classification:	Old Indo-Aryan (1500 BCE – 600 BCE)					
Descriptive Source(s):	Whitney (1993 [1889]); Cardona (2003)					
Consonant Phonemes:	p	t	ʈ	ʧ	k	
	pʰ	tʰ	ʈʰ	ʧʰ	kʰ	
	b	d	ɖ	ʤ	g	
	bʰ	dʰ	ɖʰ	ʤʰ	gʰ	
		s	ʂ	ʃ		(h) ɦ
	m	n	ɳ	ɲ	ŋ	
		l				
		r				
	v			j		
Lexical Data Source:	Apte (1957–1959)					
Search Domain:	headwords containing $\#C_1V(N)C_2$ sequences					
Segment Class 1:	$\{C_1, C_2\}$ = coronal plosives and retroflex sonorants					
Segment Class 2:	$\{C_1, C_2\}$ = coronal fricatives					

Table B-88 Sanskrit: Observed values for co-occurring coronal plosives and retroflex sonorants (n = 393)

C_1\ C_2	t	tʰ	d	dʰ	ʈ	ʈʰ	ɖ	ɖʰ	ɳ
t	80	9	39	-	14	-	43	1	15
tʰ	2	2	-	-	-	-	2	-	-
d	41	-	27	24	-	-	28	7	1
dʰ	28	-	1	4	5	-	1	-	3
ʈ	-	-	-	-	6	-	-	-	-
ʈʰ	-	-	-	-	-	1	-	-	-
ɖ	-	1	1	-	-	-	5	-	-
ɖʰ	-	-	-	-	-	-	-	2	-
ɳ	-	-	-	-	-	-	-	-	-

Table B-89 Sanskrit: Observed (O) and Expected (E) values with O/E ratios for co-occurring coronal plosives and retroflex sonorants; segments classified by place and manner (n = 393)

C_1\ C_2		t, t^h, d, d^h	ʈ, $ʈ^h$, ɖ, $ɖ^h$	ɳ
t, t^h, d, d^h	O	257	101	19
	E	248.5	110.3	18.2
	O/E	**1.03**	**0.92**	**1.04**
ʈ, $ʈ^h$, ɖ, $ɖ^h$	O	2	14	0
	E	10.5	4.7	0.8
	O/E	**0.19**	**2.99**	**0.00**

Table B-90 Sanskrit: Observed (O) and Expected (E) values with O/E ratios for co-occurring coronal fricatives (n = 322)

C_1\ C_2		s	ʂ	ʃ
s	O	124	27	61
	E	104.7	48.7	58.6
	O/E	**1.18**	**0.55**	**1.04**
ʂ	O	0	7	0
	E	3.5	1.6	1.9
	O/E	**0.00**	**4.35**	**0.00**
ʃ	O	35	40	28
	E	50.9	23.7	28.5
	O/E	**0.69**	**1.69**	**0.98**

Santali

Classification:	Munda, North					
Descriptive Source(s):	Ghosh (2008); cf. Neukom (2001); Sebeok (1943)					
Consonant Phonemes:[19]	p	t	ʈ	ʧ	k	
	(p^h)	(t^h)	(ʈ^h)	(ʧ^h)	(k^h)	
	b	d	ɖ	ʤ	ɡ	
	(b^h)	(d^h)	(ɖ^h)	(ʤ^h)	(ɡ^h)	
		s				h
	m	n		ɲ	ŋ	
		r	ɽ			
			(ɽ^h)			
		l				
	w			j		
Lexical Data Source:	Bodding (1929–1936)					
Search Domain:	headwords containing #$C_1V(N)C_2$ sequences[20]					
Segment Class:	$\{C_1, C_2\}$ = coronal plosives and retroflex sonorants					

Table B-91 Santali: Observed values for each $C_1 \ldots C_2$ pair ($n = 1315$)

$C_1 \backslash C_2$	t	t^h	d	d^h	ʈ	ʈ^h	ɖ	ɖ^h	ɽ	ɽ^h
t	67	9	18	-	-	-	-	-	98	-
t^h	36	16	-	-	-	-	-	-	8	-
d	40	-	110	25	5	1	1	-	129	9
d^h	37	-	58	4	4	1	2	-	108	-
ʈ	4	-	-	-	62	10	46	3	28	1
ʈ^h	2	-	-	-	48	16	10	-	28	-
ɖ	3	-	-	-	24	6	89	33	11	3
ɖ^h	-	-	-	-	8	6	53	4	31	-
ɽ	-	-	-	-	-	-	-	-	-	-
ɽ^h	-	-	-	-	-	-	-	-	-	-

[19] Aspirated stops occur primarily in Indo-Aryan loanwords. The few cases of aspiration in native vocabulary can be attributed to contraction (e.g., /dihiri/ > /dhiri/ 'stone').

[20] Headwords were excluded if a space intervenes between C_1 and C_2 in Bodding's transcription. Such words are morphogically complex, often the product of reduplication (e.g., /do do/ 'suck, suckle').

Table B-92 Santali: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner (n = 1315)

C_1\ C_2		t, tʰ, d, dʰ	t̪, t̪ʰ, d̪, d̪ʰ	ʈ, ʈʰ
t, tʰ, d, dʰ	O	420	14	352
	E	256.4	258.2	271.4
	O/E	1.64	0.05	1.30
t̪, t̪ʰ, d̪, d̪ʰ	O	9	418	102
	E	172.6	173.8	182.6
	O/E	0.05	2.41	0.56

Shina, Gilgit

Classification: Indo-Aryan, Northwest, Dardic

Descriptive Source(s): Radloff (1999)

Consonant Phonemes:[21]

p	t	ʈ		k	
pʰ	tʰ	ʈʰ		kʰ	
b	d	ɖ		g	
	ʦ	ʈʂ	ʧ		
	ʦʰ	ʈʂʰ	ʧʰ		
			ʤ		
	s	ʂ	ʃ		h
	z	ʐ	(ʒ)		
m	n	ɳ		ŋ	
	r	ɽ			
	l				
w			j		

Lexical Data Sources: Carla Radloff (personal communication, Fieldworks database) supplemented with additional data from Bailey (1924)

Search Domain: headwords containing $\#C_1V(N)C_2$ sequences

Segment Class: $\{C_1, C_2\}$ = coronal obstruents (plosives, affricates, fricatives)

[21] Bailey's (1924) data includes a few instances of [ʒ], which may be an allophone of /ʤ/ (cf. Radloff, 1999).

Table B-93 Shina: Observed values for each $C_1 \ldots C_2$ pair (n=243)

C_1\C_2	t	tʰ	d	ʈ	ʈʰ	ɖ	ts	tsʰ	tʃ	tʃʰ	dʒ	ʈʂ	ʈʂʰ	s	z	ʃ	ʒ	ʂ	ʐ
t	11	-	1	-	-	-	-	-	2	-	1	1	-	4	1	3	1	1	-
tʰ	-	1	-	1	-	-	-	-	-	-	-	3	-	1	-	1	-	1	-
d	2	-	6	2	-	-	-	-	-	-	3	-	2	6	5	3	1	1	-
ʈ	-	-	-	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ʈʰ	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
ɖ	-	-	-	2	-	5	-	-	-	-	-	-	-	-	-	-	-	-	-
ts	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
tsʰ	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
tʃ	1	-	5	9	-	-	-	-	3	-	1	-	-	2	-	1	-	1	-
tʃʰ	6	-	2	6	-	-	-	-	2	-	2	-	-	-	1	1	-	2	-
dʒ	3	-	1	3	-	1	-	-	-	-	1	-	-	1	1	-	-	-	4
ʈʂ	-	-	-	7	-	-	-	-	-	-	-	3	-	-	-	-	-	-	-
ʈʂʰ	2	-	-	-	-	-	-	-	-	-	-	1	1	-	-	-	-	-	-
s	10	-	6	2	-	-	-	-	2	-	4	8	-	3	4	-	-	-	-
z	3	-	3	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-
ʃ	7	-	5	3	-	-	-	-	1	-	3	-	-	-	2	6	-	1	-
ʒ	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-
ʂ	1	-	3	2	-	-	-	-	1	-	-	-	-	-	-	-	-	5	-
ʐ	-	-	-	-	-	-	-	-	-	-	-	1	-	2	-	-	-	-	-

Table B-94 Shina: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner; linear order of segments collapsed (n=243)

C₁\ C₂		t, tʰ, d	ʈ, ʈʰ, ɖ	ʦ, ʦʰ	ʧ, ʧʰ, ʤ	ʈʂ, ʈʂʰ	s, z	ʃ, ʒ	ʂ, ʐ
t, tʰ, d	O	21	3	0	24	8	39	21	7
	E	21.1	16.4	0.7	26.5	10.2	24.9	14.0	9.2
	O/E	0.99	0.18	0.00	0.91	0.79	1.57	1.50	0.76
ʈ, ʈʰ, ɖ	O		10	2	19	7	3	3	2
	E		2.2	0.4	13.0	3.7	10.9	6.5	3.7
	O/E		4.60	5.06	1.46	1.88	0.27	0.46	0.54
ʦ, ʦʰ	O			0	0	0	0	0	0
	E			0.0	0.2	0.2	0.3	0.1	0.1
	O/E			0.00	0.00	0.00	0.00	0.00	0.00
ʧ, ʧʰ, ʤ	O				9	0	11	6	8
	E				6.4	6.7	13.9	7.3	5.6
	O/E				1.40	0.00	0.79	0.82	1.44
ʈʂ, ʈʂʰ	O					5	8	1	1
	E					1.2	6.1	3.5	2.2
	O/E					4.13	1.30	0.29	0.45
s, z	O						9	2	2
	E						7.0	7.6	5.3
	O/E						1.29	0.26	0.38
ʃ, ʒ	O							6	1
	E							2.0	3.0
	O/E							2.96	0.34
ʂ, ʐ	O								5
	E								1.0
	O/E								5.06

Table B-95 Shina: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and sibilant/non-sibilant manner; linear order collapsed (n = 243)

C1\ C2		t, tʰ, d	ʈ, ʈʰ, ɖ	ʦ, ʦʰ, s, z	ʧ, ʧʰ, ʤ, ʃ, ʒ	ʈʂ, ʈʂʰ, ʂ, ʐ
t, tʰ, d	O	21	3	39	45	15
	E	21.1	16.4	25.6	40.4	19.3
	O/E	0.99	0.18	1.53	1.11	0.78
ʈ, ʈʰ, ɖ	O		10	5	22	9
	E		2.2	11.3	19.5	7.4
	O/E		4.60	0.44	1.13	1.22
ʦ, ʦʰ, s, z	O			9	13	10
	E			7.3	21.9	11.8
	O/E			1.24	0.59	0.85
ʧ, ʧʰ, ʤ, ʃ, ʒ	O				21	10
	E				15.7	18.7
	O/E				1.33	0.54
ʈʂ, ʈʂʰ, ʂ, ʐ	O					11
	E					4.4
	O/E					2.49

Table B-96 Shina: Observed (O) and Expected (E) values with O/E ratios; segments classified as retroflex/non-retroflex and sibilant/non-sibilant; linear order of segments collapsed (n = 243)

C1\ C2		t, tʰ, d	ʈ, ʈʰ, ɖ	ʦ, ʦʰ, s, z ʧ, ʧʰ, ʤ, ʃ, ʒ	ʈʂ, ʈʂʰ, ʂ, ʐ
t, tʰ, d	O	21	3	84	15
	E	21.1	16.4	66.0	19.3
	O/E	0.99	0.18	1.27	0.78
ʈ, ʈʰ, ɖ	O		10	27	9
	E		2.2	30.8	7.4
	O/E		4.60	0.88	1.22
ʦ, ʦʰ, s, z ʧ, ʧʰ, ʤ, ʃ, ʒ	O			43	20
	E			44.9	30.4
	O/E			0.96	0.66
ʈʂ, ʈʂʰ, ʂ, ʐ	O				11
	E				4.4
	O/E				2.49

Sindhi

Classification: Indo-Aryan, Northwestern

Descriptive Source(s): Khubchandani (2003); Nihalani (1999)

Consonant Phonemes:[22]

p	t		ʈ	tʃ	k	(q)	
pʰ	tʰ		ʈʰ	tʃʰ	kʰ		
b	d		ɖ	dʒ	g		
bʰ	dʰ		ɖʰ	dʒʰ	gʰ		
ɓ		ɗ		ʄ	ɠ		
f		s		ʃ	x		
		z			ɣ		h
m, (mʰ)		n, (nʰ)	ɳ, (ɳʰ)	ɲ	ŋ		
		r	ɽ, (ɽʰ)				
	l, (lʰ)						
ʋ				j			

Lexical Data Source: Turner (1969)

Search Domain: headwords containing $\#C_1V(N)C_2$ sequences

Segment Class: $\{C_1, C_2\}$ = coronal plosives, coronal implosives and retroflex sonorants

Table B-97 Sindhi: Observed values for each $C_1 \ldots C_2$ pair ($n = 183$)

$C_1 \backslash C_2$	t	tʰ	d	dʰ	ʈ	ʈʰ	ɖ	ɖʰ	ɗ	ɳ	ɽ	ɽʰ
t	4	2	2	-	-	1	-	-	3	4	12	-
tʰ	-	1	-	3	-	-	-	-	3	6	4	-
d	-	-	-	-	-	-	-	-	-	-	-	-
dʰ	11	-	-	3	-	-	-	-	-	15	13	-
ʈ	-	-	-	-	5	1	2	-	2	2	1	-
ʈʰ	-	-	-	-	-	8	-	-	1	1	-	-
ɖ	-	-	-	-	-	-	-	-	-	-	-	-
ɖʰ	-	-	-	-	-	-	-	4	-	1	-	-
ɗ	2	-	11	8	4	7	7	7	6	11	-	5
ɳ	-	-	-	-	-	-	-	-	-	-	-	-
ɽ	-	-	-	-	-	-	-	-	-	-	-	-
ɽʰ	-	-	-	-	-	-	-	-	-	-	-	-

[22] Nihalani (1999) lists aspirated (breathy voiced) sonorants /mʰ/, /nʰ/, /ɳʰ/, /ɽʰ/, /lʰ/, which are absent in Khubchandani (2003), and has retroflex /ʂ/ in place of palatal /ʃ/.

Table B-98 Sindhi: Observed (O) and Expected (E) values with O/E ratios for co-occurring coronal plosives and retroflex sonorants; segments classified by place and manner (n = 106)

C_1\ C_2		t, t^{h}, d, d^{h}	ʈ, ʈ^{h}, ɖ, ɖ^{h}	ɳ, ɽ, ɽ^{h}
t, t^{h}, d, d^{h}	O	26	1	54
	E	19.9	16.0	45.1
	O/E	1.31	0.06	1.20
ʈ, ʈ^{h}, ɖ, ɖ^{h}	O	0	20	5
	E	6.1	5.0	13.9
	O/E	0.00	4.04	0.36

Sinhala

Classification:	Indo-Aryan, Sinhalese-Maldivian					
Descriptive Source(s):	Gair (2003)					
Consonant Phonemes:[23]	p	t	ʈ	ʧ	k	
	b	d	ɖ	ʤ	g	
	^{m}b	^{n}d	^{n}ɖ	^{n}ʤ	^{n}g	
	(f)	s		(ʃ)		h
	m	n	(ɳ)	ɲ	ŋ	
		r				
		l	(ɭ)			
	w			j		
Lexical Data Source:	Turner (1969)					
Search Domain:	headwords containing #C_1V(N)C_2 sequences					
Segment Class:	{C_1, C_2} = coronal plosives and retroflex sonorants					

[23] /ɳ/ and /ɭ/ are distinguished orthographically but have merged with /n/ and /l/, respectively, in contemporary speech (Gair, 2003).

Table B-99 Sinhala: Observed values for each $C_1 \ldots C_2$ pair ($n = 90$)

C_1\ C_2	t	d	ⁿd	ʈ	ɖ	ⁿɖ	ɳ	ɭ
t	14	4	-	7	1	-	3	5
d	10	15	-	11	8	-	5	7
ⁿd	-	-	-	-	-	-	-	-
ʈ	-	-	-	-	-	-	-	-
ɖ	-	-	-	-	-	-	-	-
ⁿɖ	-	-	-	-	-	-	-	-
ɳ	-	-	-	-	-	-	-	-
ɭ	-	-	-	-	-	-	-	-

Table B-100 Sinhala: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner; laryngeal distinctions and pre-nasalization ignored ($n = 90$)

C_1\ C_2		t, d, ⁿd	ʈ, ɖ, ⁿɖ	ɳ, ɭ
t, d, ⁿd	O	43	27	20
	E	43.0	27.0	20.0
	O/E	1.00	1.00	1.00
ʈ, ɖ, ⁿɖ	O	0	0	0
	E	0.0	0.0	0.0
	O/E	0.00	0.00	0.00

Tamil

Classification:	Dravidian, South				
Descriptive Source(s):	Annamalai & Steever (1998); Keane (2004); Rajaram (1972)				
Consonant Phonemes:[24]	p	t	ʈ	ʧ	k
	(b)	(d)	(ɖ)	(ʤ)	(g)
		s			
	m	n	ɳ	(ɲ)	
		r			
		l	ɭ		
	ʋ		ɻ	j	

[24] Voiced stops are subphonemic in native vocabulary. The Kanniyakumari dialect has a distinct apico-alveolar stop, which occurs as a voiceless geminate [t̺t̺], and an apical flap/trill [r̺], which is distinct from and more retracted than /r/ (Christdas, 1988).

Lexical Data Source: Fabricius (1972)
Search Domain: headwords containing $\#C_1V(N)C_2$ sequences
Segment Class: $\{C_1, C_2\}$ = coronal plosives and retroflex sonorants

Table B-101 Tamil: Observed values for each $C_1 \ldots C_2$ pair (n = 612)

$C_1 \backslash C_2$	t	ʈ	ɳ	ɭ	ɻ
t	124	254	68	106	56
ʈ	-	2	1	-	1
ɳ	-	-	-	-	-
ɭ	-	-	-	-	-
ɻ	-	-	-	-	-

Table B-102 Tamil: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner (n = 612)

$C_1 \backslash C_2$		t	ʈ	ɳ, ɭ, ɻ
	O	124	254	230
t	E	123.2	254.3	230.5
	O/E	1.01	1.00	1.00
	O	0	2	2
ʈ	E	0.8	1.7	1.5
	O/E	0.00	1.20	1.32

Telugu

Classification: Dravidian, South-Central
Descriptive Source(s): Krishnamurti (1998); Sastry (1972)
Consonant Phonemes:

p	t	ʈ	ʧ	k	
(pʰ)	(tʰ)	(ʈʰ)	(ʧʰ)	(kʰ)	
b	d	ɖ	ʤ	g	
(bʰ)	(dʰ)	(ɖʰ)	(ʤʰ)	(gʰ)	
(f)	s	(ʂ)	(ʃ)		h
m	n	ɳ			
	r				
	l	ɭ			
w			j		

Lexical Data Source: Gwynn (1991)
Search Domain: headwords containing $\#C_1V(N)C_2$ sequences
Segment Class: $\{C_1, C_2\}$ = coronal plosives and retroflex sonorants

Table B-103 Telugu: Observed values for each $C_1 \ldots C_2$ pair ($n = 519$)

$C_1 \backslash C_2$	t	t^h	d	d^h	ʈ	ʈ^h	ɖ	ɖ^h	ɳ	ɭ
t	61	6	35	6	101	-	88	-	2	23
t^h	-	-	-	-	-	-	-	-	-	-
d	30	-	30	-	28	-	74	-	3	10
d^h	8	-	-	-	2	-	-	-	3	2
ʈ	-	-	-	-	1	-	-	-	-	-
ʈ^h	-	-	-	-	-	1	-	-	-	1
ɖ	-	-	1	-	-	-	3	-	-	-
ɖ^h	-	-	-	-	-	-	-	-	-	-
ɳ	-	-	-	-	-	-	-	-	-	-
ɭ	-	-	-	-	-	-	-	-	-	-

Table B-104 Telugu: Observed (O) and Expected (E) values with O/E ratios; segments classified by place and manner ($n = 519$)

$C_1 \backslash C_2$		t, t^h, d, d^h	ʈ, ʈ^h, ɖ, ɖ^h	ɳ, ɭ
t, t^h, d, d^h	O	176	293	43
	E	174.6	294.0	43.4
	O/E	1.01	1.00	0.99
ʈ, ʈ^h, ɖ, ɖ^h	O	1	5	1
	E	2.4	4.0	0.59
	O/E	0.42	1.24	1.69

Appendix C
The natural classes similarity metric applied to Kalasha

This appendix describes the results of applying the natural classes similarity metric of Frisch, Pierrehumbert, & Broe (2004) to the coronal consonant inventory of Kalasha. In the natural classes similarity metric, the similarity of any pair of segments is calculated as the number of shared natural classes divided by the number of shared and non-shared natural classes, as shown in (1). This metric returns a value ranging between 0 and 1, in which 1 represents the highest degree of similarity (i.e., identity) and 0 represents the lowest possible degree of similarity (i.e., the segments do not share any natural classes).

(1) Natural classes similarity metric (Frisch, Pierrehumbert, & Broe, 2004)

$$\text{similarity} = \frac{\text{shared natural classes}}{\text{shared natural classes} + \text{non-shared natural classes}}$$

Natural classes are defined in terms of phonological features. For the purpose of applying the metric to Kalasha coronals, I assume the features listed in Table C-1.

Table C-1 Features assumed for Kalasha coronal consonants

	t	tʰ	d	dʰ	ʈ	ʈʰ	ɖ	ɖʰ	ʦ	ʦʰ	ʥ	ʧ	ʧʰ	ʤ	ʤʰ	ʈʂ	ʈʂʰ	ɖʐ	s	z	ʃ	ʒ	ʂ	ʐ	n	l	ɫ	r
[cor]	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
[ant]	+	+	+	+	–	–	–	–	+	+	+	–	–	–	–	–	–	–	+	+	–	–	–	–	+	+	+	+
[dist]	+	+	+	+	–	–	–	–	+	+	+	+	+	+	+	–	–	–	+	+	+	+	–	–	+	+	–	–
[son]	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	+	+	+
[cont]	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	+	+	+	+	+	–	+	+	+
[strid]	–	–	–	–	–	–	–	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+				
[nas]	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–							+			
[lat]																										+	+	–
[voice]	–	–	+	+	–	–	+	+	–	–	+	–	–	+	+	–	–	+	–	+	–	+	–	+	+	+	+	+
[s.g.]	–	+	–	+	–	+	–	+	–	+	–	–	+	–	+	–	+	–										

The features in Table C-1 are the same as those assumed by Frisch et. al. (2004: 201) for Arabic, with only minor modifications to account for differences in the Kalasha inventory. The modifications are as follows:

- The feature [±acute], which is used to distinguish emphatic and non-emphatic coronal obstruents in Arabic, is not required for Kalasha. Other features not required for Kalasha include [pharyngeal], [radical], and [constricted glottis], which are required for the wide range of uvular, pharyngeal and glottal consonants in Arabic.
- The feature [±distributed], which is not used by Frisch et. al., is introduced to account for the contrast between laminal and apical coronals in Kalasha, most notably the contrast between (laminal) palatals and (apical) retroflexes. I assume that the distinction between /l/ and /ɫ/ can also be subsumed under this feature. According to Heegård & Mørch (2004) /l/ is laminal and /ɫ/ is apical (and frequently velarized).
- The feature [±strident], which is applied only to coronal fricatives in Arabic, is extended to all coronal obstruents in Kalasha in order to account for the distinction between [+strident] affricates and [–strident] plosives.
- The feature [+spread glottis], which is applied only to Arabic /h/, is extended to all aspirated plosives and affricates in Kalasha, while the feature [–spread glottis] (not used for Arabic) is applied to their unaspirated counterparts.

In the interest of following Frisch et. al. as closely as possible I also assume the following:

- The palatal glide /j/ is not counted among the coronal consonants. Frisch et. al. treat it as [+dorsal, –back] for Arabic.
- The following features cross-classify the entire coronal inventory: [±sonorant], [±continuant], [±anterior], and [±voice]. To these I have added [±distributed].

- The following features sub-classify limited sets: [±nasal] sub-classifies stops (i.e., all segments that are [–continuant] except Arabic /ʔ/); [±lateral] sub-classifies liquids; [±strident] sub-classifies fricatives in Arabic (but is extended to all obstruents in Kalasha for reasons noted above); [±spread glottis] sub-classifies glottals in Arabic (but is extended to all stops in Kalasha for reasons noted above).

The feature specifications in Table C-1 yield a total of 203 unique classes for the inventory of Kalasha coronal consonants. Assuming these classes, the similarity metric was applied to the inventory of Kalasha coronals using Albright's (2006) segmental similarity calculator, a Perl script designed to implement the natural classes similarity metric. The results of the computation are displayed in Table C-2.

The natural classes metric makes undesirable predictions about similarity effects in Kalasha. According to the figures in Table C-2, dental /t/ is more similar to palatal /ʧ/ (0.25) than to any retroflex stop, including /ʈ/ (0.24), /ʈʰ/ (0.15), /ɖ/ (0.12), and /ɖʰ/ (0.08). Thus, the metric predicts that dental /t/ is more likely to harmonize with palatal /ʧ/ than with retroflex stops. Contrary to this prediction, dental stops do not assimilate to palatal affricates in Kalasha, though they do assimilate to retroflex stops, regardless of laryngeal distinctions. The source of the problem lies in the fact that each feature contributes equally to the evaluation of similarity in the metric. The Kalasha data suggest that [±strident], which distinguishes affricates from unaffricated stops, contributes more to the evaluation of similarity than other features. That is, dental stops are considered more similar to retroflex stops than to any [+strident] segment, presumably because dental and retroflex stops are both [−strident].

Table C-2 Similarity of Kalasha coronals according to the natural classes metric of Frisch et. al. (2004)

	t	tʰ	d	dʰ	ʈ	ʈʰ	ɖ	ɖʰ	ʦ	ʦʰ	ʣ	ʧ	ʧʰ	ʤ	ʤʰ	ʈʂ	ʈʂʰ	ɖʐ	s	z	ʃ	ʒ	ʂ	ʐ	n	l	ɫ	r
t	1.00																											
tʰ	0.53	1.00																										
d	0.38	0.26	1.00																									
dʰ	0.26	0.38	0.61	1.00																								
ʈ	0.24	0.15	0.11	0.08	1.00																							
ʈʰ	0.15	0.24	0.08	0.11	0.49	1.00																						
ɖ	0.12	0.08	0.24	0.16	0.36	0.22	1.00																					
ɖʰ	0.08	0.12	0.16	0.25	0.22	0.36	0.54	1.00																				
ʦ	0.49	0.33	0.24	0.19	0.13	0.09	0.07	0.06	1.00																			
ʦʰ	0.34	0.50	0.19	0.25	0.09	0.13	0.06	0.07	0.57	1.00																		
ʣ	0.24	0.19	0.56	0.42	0.07	0.05	0.14	0.11	0.37	0.26	1.00																	
ʧ	0.25	0.18	0.13	0.10	0.23	0.15	0.11	0.08	0.46	0.29	0.20	1.00																
ʧʰ	0.18	0.25	0.10	0.13	0.15	0.23	0.08	0.12	0.29	0.47	0.14	0.53	1.00															
ʤ	0.14	0.11	0.29	0.22	0.11	0.08	0.24	0.17	0.21	0.15	0.48	0.35	0.23	1.00														
ʤʰ	0.11	0.14	0.23	0.30	0.09	0.12	0.18	0.26	0.15	0.22	0.33	0.24	0.37	0.59	1.00													
ʈʂ	0.13	0.09	0.07	0.05	0.45	0.28	0.20	0.15	0.21	0.14	0.10	0.44	0.27	0.19	0.13	1.00												
ʈʂʰ	0.09	0.13	0.05	0.07	0.29	0.46	0.15	0.21	0.14	0.21	0.07	0.27	0.45	0.13	0.20	0.54	1.00											
ɖʐ	0.07	0.06	0.14	0.11	0.21	0.15	0.49	0.33	0.11	0.08	0.22	0.19	0.13	0.44	0.29	0.36	0.23	1.00										
s	0.20	0.20	0.12	0.12	0.06	0.06	0.04	0.04	0.30	0.31	0.15	0.16	0.16	0.09	0.09	0.08	0.08	0.05	1.00									
z	0.12	0.12	0.23	0.24	0.03	0.03	0.06	0.07	0.15	0.15	0.32	0.08	0.08	0.17	0.17	0.04	0.04	0.09	0.40	1.00								
ʃ	0.11	0.11	0.06	0.07	0.10	0.10	0.05	0.06	0.17	0.17	0.09	0.29	0.29	0.13	0.14	0.16	0.16	0.08	0.44	0.20	1.00							
ʒ	0.07	0.07	0.13	0.13	0.05	0.05	0.11	0.11	0.09	0.09	0.18	0.13	0.13	0.29	0.31	0.08	0.08	0.16	0.22	0.45	0.37	1.00						
ʂ	0.06	0.06	0.03	0.03	0.19	0.19	0.10	0.10	0.08	0.08	0.04	0.15	0.15	0.07	0.08	0.30	0.30	0.15	0.20	0.10	0.41	0.18	1.00					
ʐ	0.04	0.04	0.06	0.07	0.10	0.10	0.21	0.21	0.05	0.05	0.09	0.07	0.07	0.15	0.16	0.14	0.14	0.31	0.11	0.20	0.18	0.40	0.37	1.00				
n	0.16	0.16	0.33	0.33	0.05	0.05	0.09	0.09	0.13	0.13	0.27	0.07	0.07	0.14	0.14	0.04	0.04	0.07	0.10	0.19	0.05	0.09	0.03	0.05	1.00			
l	0.08	0.08	0.15	0.15	0.02	0.02	0.04	0.04	0.07	0.07	0.13	0.03	0.03	0.06	0.06	0.02	0.02	0.03	0.21	0.41	0.09	0.19	0.05	0.09	0.36	1.00		
ɫ	0.04	0.04	0.08	0.08	0.05	0.05	0.09	0.09	0.03	0.04	0.06	0.02	0.02	0.03	0.03	0.04	0.04	0.07	0.10	0.19	0.05	0.09	0.11	0.23	0.17	0.41	1.00	
r	0.04	0.04	0.08	0.08	0.05	0.05	0.09	0.09	0.04	0.04	0.07	0.02	0.02	0.03	0.03	0.04	0.04	0.08	0.11	0.19	0.05	0.10	0.12	0.24	0.18	0.37	0.83	1.00

Some studies have pointed out that the natural classes similarity metric can make erroneous predictions when applied to asymmetric phoneme inventories (Hansson, 2001, pp. 435-436; 2010, pp. 330-331; Mackenzie, 2009, pp. 63-64). The Kalasha coronal inventory does contain some asymmetries. First of all, dental and retroflex obstruents distinguish three manners of articulation (e.g., /t, ʦ, s/ and /ʈ, ʈʂ, ʂ/), whereas palatal obstruents distinguish only two (e.g., /ʧ, ʃ/). The palatal obstruents lack non-affricated stop counterparts. As a result, palatal obstruents have fewer distinct natural classes than dental and retroflex obstruents. This has the potential to lower the denominator (specifically, the number of non-shared natural classes) in any equation involving palatals, thereby creating higher similarity scores for palatals over retroflexes. Secondly, there is a minor asymmetry involving laryngeal features. Palatal affricates have a four-way laryngeal contrast that includes breathy voiced aspiration (i.e., /ʧ, ʧʰ, ʤ, ʤʰ/), whereas dental and retroflex affricates have a three-way contrast that lacks breathy voiced aspiration (i.e., /ʦ, ʦʰ, ʣ/ and /ʈʂ, ʈʂʰ, ɖʐ/).

In order to determine whether these asymmetries are responsible for the undesirable predictions of the metric, the calculations were run again over a hypothetical inventory, Kalasha′, which is exactly like Kalasha except that it has a non-affricated palatal stop /c/ in addition to /ʧ, ʃ/ and lacks laryngeal contrasts altogether. The laryngeal asymmetry was eliminated by ignoring laryngeal features altogether (as opposed to positing /ʣʰ/ and /ɖʐʰ/) because laryngeal features do not appear to play any role in the observed similarity effects in Kalasha. The results of applying the metric to hypothetical Kalasha′ are shown in Table C-3.

Table C-3 Similarity of coronals in hypothetical symmetric Kalasha′

	t	c	ʈ	ts	tʃ	ʈʂ	s	ʃ	ʂ	n	l	ɫ	r
t	1.00												
c	0.48	1.00											
ʈ	0.25	0.43	1.00										
ts	0.61	0.32	0.18	1.00									
tʃ	0.31	0.52	0.26	0.46	1.00								
ʈʂ	0.18	0.27	0.53	0.24	0.42	1.00							
s	0.30	0.17	0.11	0.41	0.22	0.13	1.00						
ʃ	0.17	0.26	0.15	0.23	0.41	0.22	0.46	1.00					
ʂ	0.11	0.15	0.29	0.13	0.21	0.43	0.23	0.41	1.00				
n	0.40	0.22	0.14	0.35	0.19	0.13	0.21	0.12	0.09	1.00			
l	0.19	0.11	0.08	0.17	0.09	0.07	0.39	0.19	0.12	0.39	1.00		
ɫ	0.12	0.07	0.14	0.10	0.06	0.12	0.20	0.12	0.24	0.21	0.44	1.00	
r	0.12	0.08	0.14	0.11	0.07	0.13	0.21	0.12	0.25	0.22	0.39	0.77	1.00

The figures in Table C-3 reveal that the natural classes similarity metric continues to make the same erroneous prediction even when a symmetrical system is assumed. The pair /t, tʃ/ (0.31) is still deemed more similar than the pair /t, ʈ/ (0.25). Thus, contrary to fact, the metric continues to predict that /t, tʃ/ are more likely to assimilate than /t, ʈ/. This suggests that the erroneous predicitons of the metric do not stem from the asymmetries in the Kalasha inventory, but rather, from the equal weighting of all features in the evaluation of similarity.