Areal sound change and the distributional typology of affricate richness in Eurasia

Dmitry Nikolaev and Eitan Grossman
Hebrew University of Jerusalem

This paper makes a contribution to phonological typology by investigating the distribution of affricate-rich languages in Eurasia. It shows that affricate-rich and affricate-dense languages cluster areally within Eurasia and have area-specific histories. In particular, the affricate-rich areas of western Eurasia – a 'European' area and a Caucasian area – are not the result of contact-induced sound changes or borrowing, while the two affricate-rich areas of eastern Eurasia – the Hindukush area and the Eastern Himalayan area – are the result of contact. Specifically, affricate-dense areas depend on the emergence of retroflex affricates. Moreover, languages outside these affricate-dense areas tend to lose retroflex affricates.

Keywords: language typology, distributional typology, phonology, areal linguistics, areal sound change, language contact, historical phonology, consonant inventories, affricates

1. Introduction

The goal of this paper is to investigate the distribution of Eurasian languages with rich or dense inventories of affricate segments, as a probe for the distributional typology (Bickel 2015) of complex consonant inventories. Distributional typology aims to answer the question 'what's where why when?' (Bickel 2007, 2015), i.e., to account for extant linguistic diversity in the world's languages, by accounting for the historical processes that lead to observed distributions of linguistic properties in space and time. On phonological typology, see the classic study by Maddieson (1984) and, recently, Gordon (2016).

https://doi.org/10.1075/sl.17043.nik
Studies in Language 42:3 (2018), **562–599**
ISSN 0378-4177 / E-ISSN 1569-9978

Affricates[1] are a very common class of consonants (66.5% of languages in the UPSID sample have one or more of them [Maddieson & Precoda 1992], most typically /ts/), but on average they are limited in their total count in any given language: only 22.8% of languages in the UPSID have 3 or more affricates, only 15.7% have 4 or more, and only 6.4% (29) languages have 5 or more (cf. more data on the worldwide distribution of affricates below in § 2.1.1). In contrast, 39.9% of languages have 5 or more fricatives. This relative rarity of affricate-rich languages makes them an interesting topic for a distributional-typological study, as we may hope to uncover a limited set of diachronic scenarios that lead to their emergence.

There are at least two ways of measuring the richness of segment inventories with respect to a particular segment type. The first is to simply count segments in a given sound inventory. However, simple affricate counts are a rather unreliable way of establishing the consonant richness of a language, since they are easily multiplied by the phonologisation of secondary articulations. For example, a language can easily phonologise palatalised or labialised allophones of pre-existing obstruents. Furthermore, affricate inventories can be enlarged by the phonologisation of voiced or aspirated allophones. These processes can sometimes lead to significant expansions in the number of affricates in a given inventory, as it was considered to have happened during the evolution of Abkhaz-Adyghe languages (Chirikba 1999); cf. also the fluctuations in the presence of voiced stops and affricates in varieties of Chinese and Tibetan (Norman 1988; Denwood 1999).

A second, more robust measure of affricate richness, which controls for this possible effect, is to count not the affricates themselves, but rather their places of articulation, under the assumption that the existence of multiple places of articulation is diachronically more stable than the existence of multiple affricates within a single place of articulation. We will call a language *affricate-rich* if it has affricates in at least three places of articulation.[2] Moreover, we will call a language *affricate-dense* if it has affricates in at least three places of articulation in the coronal-palatal range, i.e., excluding labial, velar, and postvelar regions; the usefulness of this term will become evident later.

1. Characterised by Ladefoged and Maddieson as "stops in which the release of the constriction is modified in such a way as to produce a more prolonged period of frication after the release" (1996: 90).

2. However, we collapse dental and alveolar affricates if they are also distinguished by an additional articulation. Thus, for the Qiangic language Ersu, which opposes pure dental and alveolar affricates, we count both places of articulation, while for Lithuanian, where dental affricates are alveolar if they are palatalised, we collapse them into one.

The aim of this paper is to investigate the distribution of affricate-rich and affricate-dense languages in Eurasia based on the data contained in the Database of Eurasian Phonological Inventories, EURPhon (Nikolaev et al. 2015).[3] For the EURPhon database, Eurasia – understood as including the Atlantic islands up to Iceland and those along the eastern and southeastern coast of the main landmass, but excluding the Indonesian archipelago and the Philippines – is sampled as densely as possible. At the time of the writing of this paper, the database includes data for 445 languages. Dialects are represented sporadically in the database, but were excluded for the present analysis. The data are extracted from descriptions of individual language varieties contained in grammars, phonological analyses, and descriptive sketches published independently or as parts of overview works on different language families or regions. Data from existing databases are not reused, and no first-hand analyses of raw data (such as wordlists and text collections) are undertaken. The phonological entities logged in the database are phonemes, but see Nikolaev (2018) for a detailed discussion of the precise criteria for data analysis and collection.

In § 2, we survey the geographical distribution of affricate-rich languages and show that it has a clear and non-trivial areal structure. We also show that this structure can be used to contextualise several claims about phonological areas in Eurasia made in the literature, especially the Hindukush linguistic area hypothesis (Edelman 1980, 1983; Tikkanen 1988, 1999, 2008; Liljegren 2017; Liljegren & Svärd 2017). In § 3, we (i) survey the sound changes that were involved in the process of emergence of some of these areas, (ii) investigate the role of retroflex affricates in their emergence and development, and (iii) interpret these processes in the framework of the theory of contact-induced sound change (Blevins 2017). In § 4, we present evidence in favour of the hypothesis that affricate-rich inventories are unstable outside linguistic areas that possess certain characteristics. Section 5 concludes.

2. Affricate richness in the languages of the world and in Eurasia

In this section, we first provide an overview of the distribution of affricate-rich languages in the world and in Eurasia. We then explore the affricate inventories of Eurasian languages from different areal clusters. We show that of the four spatial clusters seen on the map in Figure 1, only two can be regarded as true linguistic areas[4]

3. http://eurasianphonology.info/

4. We use the term '(phonological) linguistic area' in the restricted sense of a geographical region where a phonological feature is shared due to lexical borrowing or contact-induced sound change. Cf. the discussion of the notion by Campbell (2017).

as far as affricate inventories are concerned. Furthermore, we demonstrate that the overall distribution of affricate-rich languages is correlated with the distribution of retroflex affricates. The distribution of affricate-rich vis-a-vis non-affricate-rich languages in Eurasia is shown on the map in Figure 7 in Appendix A.

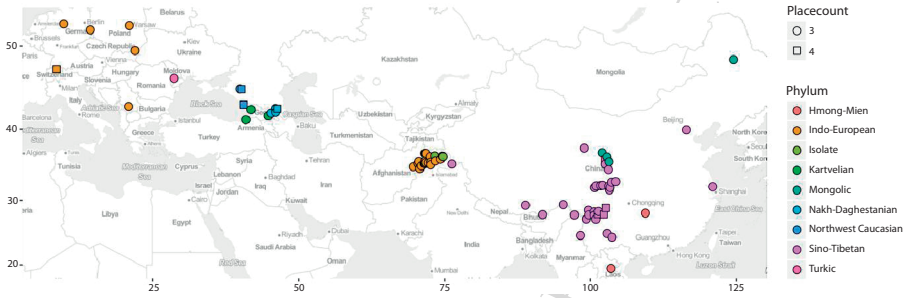

Figure 1. Distribution of affricate-rich languages in Eurasia (Nikolaev et al. 2015)

2.1 The overall distribution

2.1.1 *Affricate richness in the world's languages*

Table 2 presents the distribution of the number of places of articulation for affricates computed based on the 2222 language sample pooled from EURPhon (Nikolaev et al. 2015) and PHOIBLE (Moran et al. 2014) databases. The table shows that the modal value is 1 (nearly half of all the cases) and that the overwhelming majority of languages (93%) have less than 3 different places of articulation for affricates.

Languages having more than 3 different places of articulation for affricates are extremely rare,[5] and it is noteworthy that of 130 3-place ones nearly 66% (N = 81) are located in Eurasia, while Eurasian languages comprise only around 30% of the sample. Therefore, Eurasian languages tend to be affricate rich not only in terms of raw segment counts, but also in terms of the number of different places of articulation for affricates. Relative frequencies of the affricates found in more than 10% of the languages are shown in Table 1. It may be noted that in 88% of the cases the presence of an affricate at some place of articulation in an inventory coincides with the presence of a fricative at the same place of articulation. The converse, however, is not true: distributions of fricatives are in most cases much wider than distributions of corresponding affricates. Eurasian languages on average have more than 4 different places of articulation for fricatives, and 25% of languages have more than 5.

5. Both the 5-place ones, Banjun and Ronga, are African, while of 15 4-place ones, 2 are from the Americas, 4 are from Africa, 8 are from Eurasia, and one is from the Pacific.

Table 1. Relative frequencies of the most common affricates in the languages of Eurasia

Affricate	**tʃ**	**ts**	**dʒ**	**tsʰ**	**dz**	**tɕ**
Relative frequency	0.54	0.52	0.47	0.32	0.31	0.27
Affricate	tʃʰ	dʑ	tɕʰ	ʈʂ	ʈʂʰ	ɖʐ
Relative frequency	0.24	0.19	0.18	0.17	0.13	0.11

Table 2. Distribution of the number of different places of articulation for affricates in the languages of the world

# of places of articulation	0	1	2	3	4	5
# of languages	663	961	451	130	5	2

Sources sometimes disagree on the number of places of articulation for different languages. However, this does not have a significant impact on the analysis presented below because (i) in the vast majority of cases a later description supersedes an earlier one by providing a more precise description, and (ii) the sources never disagree on the number of different places of articulation for affricates. In older descriptions, an ambiguous notation is sometimes used (with affricates described using characters such 'c' or 'j'), but we have not encountered a case of an affricate postulated by one source being described as a stop in another one.

2.1.2 *Affricate-rich languages in Eurasia*

In the EURPhon sample used in the study, 76 out of 445 are affricate rich. Their distribution is presented in Figure 1. The languages are colour-coded according to phylum. Four areas can be discerned:

1. The European area (consisting of Standard and Zurich German, several Slavic languages, Vlax Romani, and Gagauz)
2. The Caucasian area (languages from the Abkhaz-Adyghe, Nakh-Daghestanian, and Kartvelian phyla)
3. The Hindukush area (Pamir Iranian and Dardic languages, Burushaski, and one Western Tibetic variety)
4. The Eastern Himalayan area (Khams and Amdo Tibetic, Qiangic, rGyalrongic, Qinghai-Gansu Mongolic, varieties of Mandarin Chinese, LoloBurmese, Salar, Hmong-Njua, and possibly Daur)

It will be shown that this complex distribution of a seemingly structural property – affricate richness – has a very strong and simple material correlate, namely, the presence of retroflex affricates in the inventory. First, however, we survey the individual areas.

2.2 The European area

The affricate-rich languages of the European area in our sample are listed in Table 3.[6] The two German varieties in this area – Standard German and Zurich German – are outliers: their affricate richness is due to peripheral segments (labiodental and velar affricates), absent in all other languages from this cluster. As such, they do not meet the criterion we have established for affricate density. In contrast, the other languages in this area are affricate-dense. Consequently, what we are probably dealing with here is a cluster of Slavic languages of East and South Europe, augmented by

Table 3. Affricate-rich languages of the European area

Phylum	Language	Affricate inventory
Indo-European (Germanic)	Standard German	Labiodental: /pf/
		Denti-alveolar: /ts/
		Postalveolar: /tʃ, dʒ/
	Zurich German	Labiodental: /pf/
		Denti-alveolar: /ts/
		Postalveolar: /tʃ/
		Velar: /kx/
Indo-European (Slavic)	Lower Sorbian	Denti-alveolar: /ts/
		Postalveolar: /tʃ, dʒ/
		Alveolo-palatal: /tɕ, dʑ/
	Standard Polish	Denti-alveolar: /ts, dz/
		Retroflex: /ʈʂ, ɖʐ/
		Alveolo-palatal: /tɕ, dʑ/
	Rusyn (Lemko)	Denti-alveolar: /ts, dz/
		Retroflex: /ʈʂ, ɖʐ/
		Alveolo-palatal: /tɕ, dʑ/
Indo-European (Indo-Aryan)	Vlax Romani	Denti-alveolar: /ts/
		Postalveolar: /tʃʰ, dʒ/
		Alveolo-palatal: /tɕʰ, dʑ/
Turkic	Standard Gagauz	Denti-alveolar: /ts/
		Postalveolar: /tʃ/
		Alveolo-palatal: /dʑ/

6. For a general overview of consonant inventories of European languages, cf. Stolz & Levkovych (2017). Two additional languages that may be (arguably) analyzed as affricate-dense in the European area are Skolt Sami and Hungarian (Juho Pystynen, p.c.). Mikhail Zhivlov points out (p.c.) that Standard Basque and some Basque dialects might be considered affricate dense, with three sets of affricates: dorso-alveolar (laminal), apico-alveolar, and prepalatal. However, since we treat two of these as manners at the same place of articulation, these Basque varieties are not technically affricate dense in our sense.

Vlax Romani and Gagauz. The status of Gagauz as a true affricate-rich language, moreover, is dubious since the opposition in the place of articulation between /tʃ/ and /dʒ/ is augmented by that of VOT. In other words, the only clear case of possible areal influence here is that of several Slavic languages on Vlax Romani.

2.3 The Caucasian area

The affricate-rich languages from the Caucasian area are shown in Figure 2. The language names and their affricate inventories are given in Table 4. Several observations are immediately apparent from this table. First is the simple fact of the prevalence of affricate richness in the languages of this area. Furthermore, unlike the languages in the European area (see above), many of the languages in this area have three or more affricates at the same place of articulation.

It may seem that this apparent geographical clustering lends support to a popular, albeit controversial, notion of the Caucasus as a linguistic area (for a recent overview, see Grawunder [2017]). Indeed, all these languages have numerous places of articulation for affricates and have ejective affricates, otherwise extremely rare in Eurasia (to our knowledge, they are only attested in Itelmen). At the same time, however, the picture is more nuanced, since languages from different families have very different affricate inventories.

Starting from the west, there are the Abkhaz-Adyghe languages. All of them are affricate-dense with no lateral affricates; two of them possess typologically unique hissing-hushing affricates, which are described by Ladefoged and Maddieson (1996) for the related language Ubykh as 'laminal closed postalveolar'. Kartvelian languages – whose affricate inventories are in fact identical – are in territorial proximity to the Abkhaz-Adyghe languages, but their affricate inventories are not dense and are rich only due to the postvelar affricate /qχ'/.

Postvelar affricates are also found in some of the Nakh-Daghestanian languages, whose common feature, however, is the presence of lateral affricates (Bats is known to be heavily influenced by Georgian). Thus, we see that each of the three groups of languages from different families has its own way of achieving affricate richness, and it is hard to argue that the specific affricate richness observed is a contact-derived phenomenon in this region, since there is no evident segment borrowing or contact-induced sound changes. This means that the sources of affricate richness in the Caucasus are more or less unknown, since there is at present no explanation for contact-induced preferences for large or small inventories that are indifferent to the actual segments or contrasts involved (cf. also § 2.6).

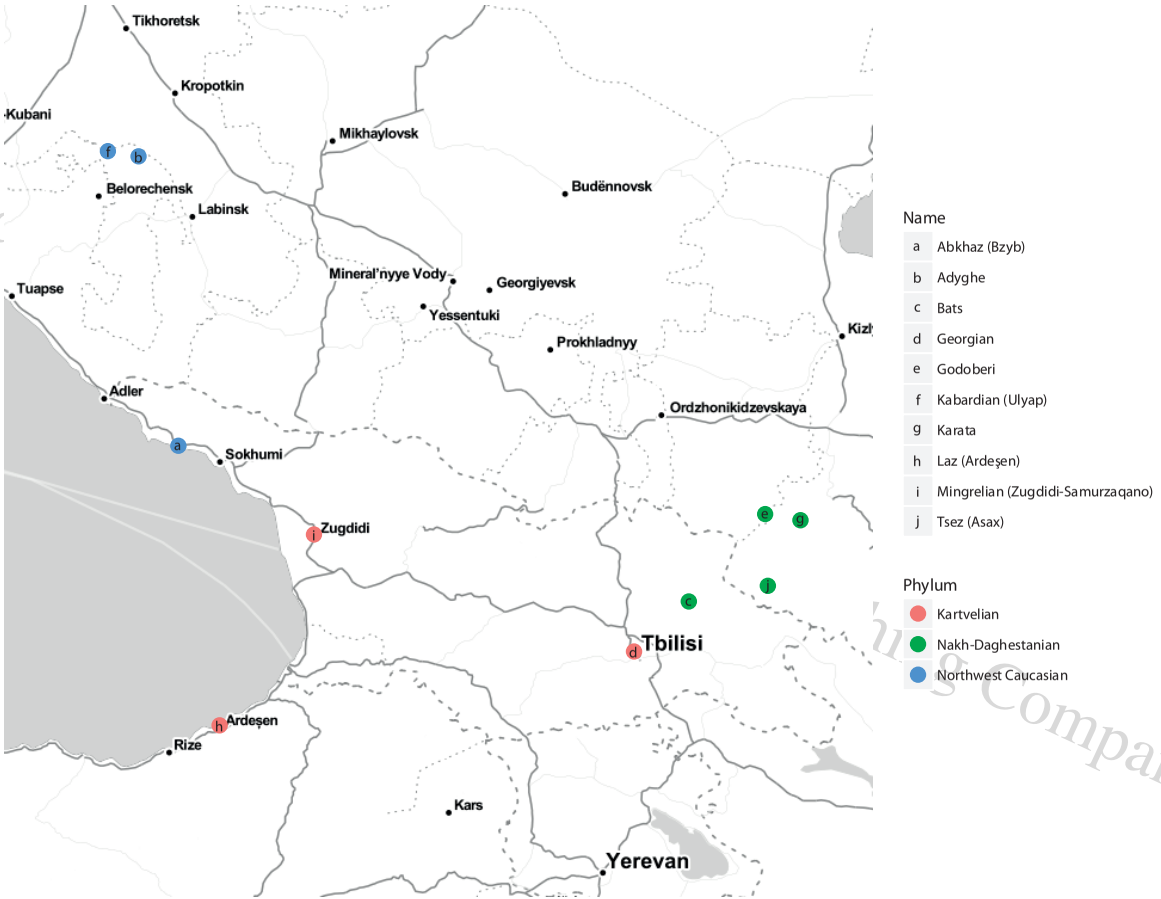

Figure 2. Affricate-rich languages in the Caucasus (Nikolaev et al. 2015)

Table 4. Affricate inventories of the languages of the Caucasian area

Phylum	Language	Affricate inventory
Abkhaz-Adyge	Kabardian[7]	Denti-alveolar: /ts, ts', ts'^{w}, dz/
		Postalveolar: /tʃ, tʃ'/
		Alveolo-palatal: /tɕ, tɕ', dʑ/
	Abkhaz	Denti-alveolar: /ts, ts', dz/
		Postalveolar: /tʃ, tʃ'/
		Alveolo-palatal: /tɕ, tɕ', dʑ/
		Hissing-hushing:[8] /tŝ^{w}, dẑ^{w}/
	Adyghe	Denti-alveolar: /ts^{h}, ts', dz/
		Postalveolar: /tʃ^{h}, tʃ', dʒ/
		Hissing-hushing: /tŝ^{h}, tŝ', tŝ^{wh}, tŝ^{w}', dẑ, dẑ^{w}/
		Retroflex: /tʂ^{h}, tʂ', dʐ/
Kartvelian	Georgian, Mingrelian, Laz	Denti-alveolar: /ts, ts', dz/
		Postalveolar: /tʃ, tʃ', dʒ/
		Uvular: /qχ'/
Nakh-Daghestanian	Godoberi	Denti-alveolar: /ts^{h}, tsː^{h}, ts'/
		Postalveolar: /tʃ^{h}, tʃː^{h}, tʃ', dʒ/
		Lateral: /tɬː^{h}, tɬ'/
		Uvular: /qχ^{h}/
	Karata	Denti-alveolar: /ts^{h}, tsː^{h}, ts', tsː'/
		Postalveolar: /tʃ^{h}, tʃː^{h}, tʃ', tʃː', dʒ/
		Lateral: /tɬː^{h}, tɬ', tɬː'/
		Velar: /kxː^{h}, kxː'/
		Uvular: /qχː', qχː^{h}/
	Tsez	Denti-alveolars: /ts, ts'/
		Postalveolar: /tʃ, tʃ'/
		Lateral: /tɬ, tɬ'/
	Bats	Denti-alveolars: /ts^{h}, ts', dz/
		Postalveolar: /tʃ^{h}, tʃ', dʒ/
		Uvular: /qχ'/

7. It may be pointed out that the dialect of Kabardian spoken in Turkey retained only the denti-alveolar affricates (Gordon & Applebaum 2006). Cf. the discussion on the loss of retroflex affricates outside specific areas below. We thank an anonymous reviewer for the reference.

8. Cf. a note on their properties in the text.

2.4 The Hindukush linguistic area

The Hindukush seems to have been the first linguistic area that was established based on the affricates contained in the phonological inventories of the languages. A representative sample of the languages from this area with their affricate inventories is given in Table 5 (the full list is given in Appendix B). The fact that languages from different genera (Iranian, Dardic, Nuristani) found in this region have retroflex affricates, while other languages from the same genera outside this region lack them prompted Edelman (1980, 1983) to hypothesise that this must be due to the areal influence of Burushaski. She also identified several lexical and grammatical features uniting the languages of what she sometimes called the 'Himalayan linguistic area' or 'Central Asiatic linguistic area'. An analysis of a fuller dataset showing the same areal structure and the prominence of affricates and especially retroflex affricates in the Hindukush area was recently presented by Liljegren (2017).

Edelman's argument was refined by Tikkanen (1988, 1999, 2008), who investigated the distribution of different linguistic features in the South Asian subcontinent and on its borders. He noted that historical and distributional data point to the northward spread of retroflex segments, which possibly originated in South India. Therefore, Burushaski is an unlikely source for retroflex affricates in the neighbouring languages. It may be noted that the only argument in favour of Edelman's theory is the fact that, unlike for Iranian, Dardic, and Nuristani languages, we may not be sure that proto-Burushaski did not have retroflex affricates, while in other languages their innovative character seems indisputable.

Unlike in the two previous cases, affricate inventories in the languages of the Hindukush are absolutely uniform: they are all dense, consisting of a dental series, a postalveolar (palato-alveolar or alveolo-palatal) series, and a retroflex series (Dardic Gawarbati, situated at the centre of the area, can serve as a representative

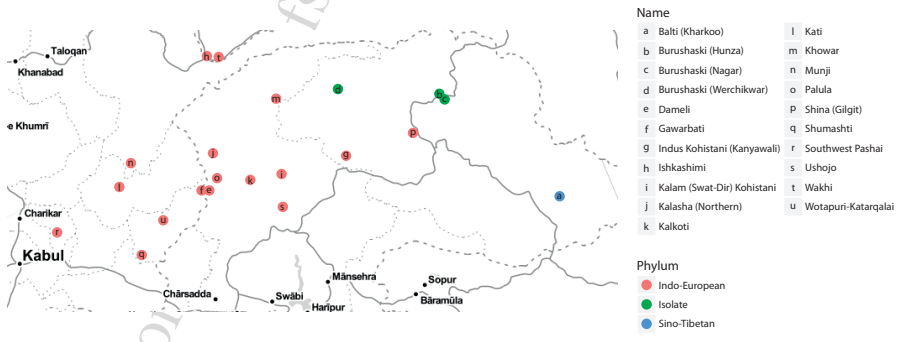

Figure 3. Affricate-rich languages in the Hindukush linguistic area (Nikolaev et al. 2015)

example, with an /ts, tsʰ/ + /tʂ, tʂʰ/ + /tʃ, dʒ/) inventory, and there can be no doubt of the contact origin for these commonalities (cf. the discussion of areally restricted sound-change processes in the next section). The main question is where to look for their origin. Edelman (1980, 1983) and Tikkanen (1988, 1999, 2008) were inclined to look for a local source, either Burushaski or some other substrate language, which would have disappeared after having influenced the mostly Indo-European languages in the area. It is impossible to disprove such a scenario, but it is tempting to find another explanation that would also account for a very similar (but hitherto unreported) affricate-rich linguistic area in the Eastern Himalayas.

Table 5. Affricate inventories of the languages of the Hindukush area (sample)

Phylum	Language	Affricate inventory
Burushaski	Hunza Burushaski	Denti-alveolar: /ts, tsʰ/
		Retroflex: /ʦ, ʦʰ, ʣ/
		Alveolo-palatal: /tɕ, tɕʰ, dʑ/
Indo-European (Iranian)	Wakhi	Denti-alveolar: /ts, dz/
		Retroflex: /ʦ, ʣ/
		Postalveolar: /tʃ, dʒ/
Indo-European (Dardic)	Khowar	Denti-alveolar: /ts, tsʰ, dz/
		Retroflex: /ʦ, ʦʰ, ʣ/
		Alveolo-palatal: /tɕ, tɕʰ, dʑ/
Indo-European (Nuristani)	Kati	Denti-alveolar: /ts, tsʲ, tsʷ, dz/
		Retroflex: /ʦ, ʦʷ, ʣ/
		Postalveolar: /tʃ, tʃʷ/
Sino-Tibetan (Tibetic)	Western Balti	Denti-alveolar: /ts, tsʰ/
		Retroflex: /ʦ, ʦʰ, ʣ/
		Postalveolar: /tʃ, tʃʰ, dʒ/

2.5 The Eastern Himalayan linguistic area

The Eastern Himalayan group of languages is the largest in the sample (both in terms of number of languages and in terms of territory),[9] but at the same time it is also remarkably uniform. A representative sample of the languages from this area and their affricate inventories is given in Table 6 (the full list of languages with their affricate inventories is given in Appendix C). Although some languages display highly unusual inventories (cf. the already-mentioned Qiangic language Ersu, which opposes dental and alveolar affricates: /t̪s̪, t̪s̪ʰ, d̪z̪, ⁿt̪s̪ʰ, ⁿd̪z̪/ + /ʦ, ʦʰ,

9. Several more northern Hmongic languages situated in the southern part of the area were reported to be affricate rich, but the ambiguities in their descriptions made their inclusion in the database undesirable.

dz, ⁿtsʰ, ⁿdz/ + /tɕ, tɕʰ, dʑ, ⁿtɕʰ, ⁿdʑ/ + /ʈɽ, ʈɽʰ, ɖɽ, ⁿʈɽʰ, ⁿɖɽ/) most of them share the 'basic' set of dental, postalveolar, and retroflex affricates, similarly to the Hindukush languages. There are exceptions (for example, Sangdam Tibetan: /ts, tsʰ, dz/ + /tɕ, tɕʰ, dʑ/ + /cç, cçʰ, ɟʝ/), but all of them are affricate-dense and none involve peripheral places of articulation.

Given the similarities between this area and the Hindukush region, it is tempting to provide a unifying scenario for the emergence of affricate-dense inventories in both areas, especially since we have reason to believe that the key ingredient for a dense affricate inventory – the retroflex affricate series – did not belong to any of the proto-languages of the region (see the discussion of sound-change processes in § 3). As such, Burushaski or another substrate language from Western Himalayas is a poor candidate for the progenitor of this areal structure as retroflex-rich inventories are conspicuously absent from the Sino-Tibetan and Indo-Aryan languages of Nepal, Bhutan, North-West India (excluding Jammu and Kashmir) and central Tibet (cf. Figure 5; no less than 35 languages from the EURPhon database fall within this area).

A possible scenario accounting for both regions simultaneously is a general diffusion of retroflex segments from South Asia, which could have led to the overlapping of the retroflex area with two distinct areas: the Hindukush area and the Eastern Himalayan area, both of which are characterised by large consonant inventories comprising several affricates, and both of which are high-altitude 'accretion' zones [Nichols 1997]).

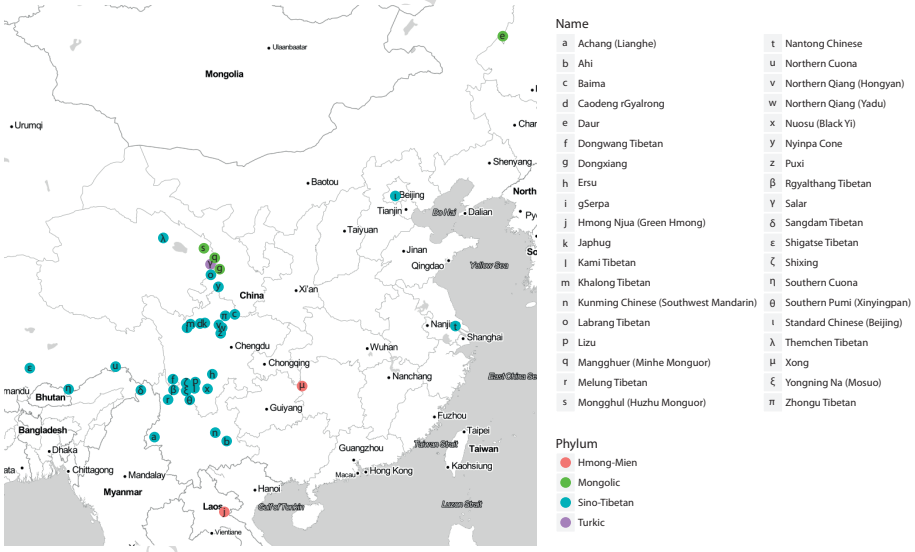

Figure 4. Affricate-rich languages in the Eastern Himalayan linguistic area (Nikolaev et al. 2015)

Table 6. Affricate inventories of the languages of the Eastern Himalayan area (sample)

Phylum	Language	Affricate inventory
Mongolic (Qinghai-Gansu)	Monggul	Denti-alveolar: /ts, tsʰ/ Retroflex: /ʈs, ʈsʰ/ Postalveolar: /tʃ, tʃʰ/
Sino-Tibetan (Qiangic)	Northern Qiang (Yadu)	Denti-alveolar: /ts, tsʰ, dz/ Retroflex: /ʈs, ʈsʰ, ɖzˌ/ Alveolo-palatal: /tɕ, tɕʰ, dʑ/
Sino-Tibetan (rGyalrongic)	Japhug	Denti-alveolar: /ts, tsʰ, dz, ⁿdz/ Retroflex: /ʈs, ʈsʰ, ɖzˌ, ⁿɖzˌ/ Alveolo-palatal: /tɕ, tɕʰ, dʑ, ⁿdʑ/
Sino-Tibetan (Lolo-Burmese)	Achang	Denti-alveolar: /ts, tsʰ/ Retroflex: /ʈs, ʈsʰ/ Alveolo-palatal: /tɕ, tɕʰ/
Hmong-Mien (Hmongic)	Xong	Denti-alveolar: /ts, tsʰ, ⁿts, ⁿtsʰ/ Retroflex: /ʈs, ʈsʰ, ⁿʈs, ⁿʈsʰ/ Alveolo-palatal: /tɕ, tɕʰ, ⁿtɕ, ⁿtɕⁿ/

2.6 Affricate-rich areas and retroflex affricates

Before discussing the role of retroflex affricates in the emergence of affricate-rich areas, it is important to point out that they rarely appear in the inventories on their own. In the pooled worldwide dataset of 2222 languages, only 3 (Eastern Khanty, Tsou, and Pohnpeian) have only retroflex affricates. Of the 113 languages with retroflex affricates, 94 also have denti-alveolar affricates, 102 also have postalveolar or alveolo-palatal affricates, and 86 (76%) have affricates at both denti-alveolar and post-alveolar places of articulation. Of this particular type of affricate-rich inventories, 70% (60 languages) are found in Eurasia. As such, it is a very robust statistical universal that the presence of retroflex affricates implies the presence of affricates at two additional coronal places of articulation.

On the other hand, out of 1261 languages with either postalveolar or alveolo-palatal affricates and out of 741 languages with denti-alveolar affricates, only 505 have both types of affricates. This shows that denti-alveolar and postalveolar affricates are much more likely to appear on their own, as compared to retroflex affricates, and therefore are not strongly linked to the emergence of affricate-rich inventories.

The crucial role that retroflex affricates play in the formation of the Hindukush and Eastern Himalayan areas – and affricate-rich areas in general – is evident from Figure 5, which shows the distribution of retroflex affricates in Eurasia. They are omnipresent in Hindukush and the Eastern Himalayas and appear both in the

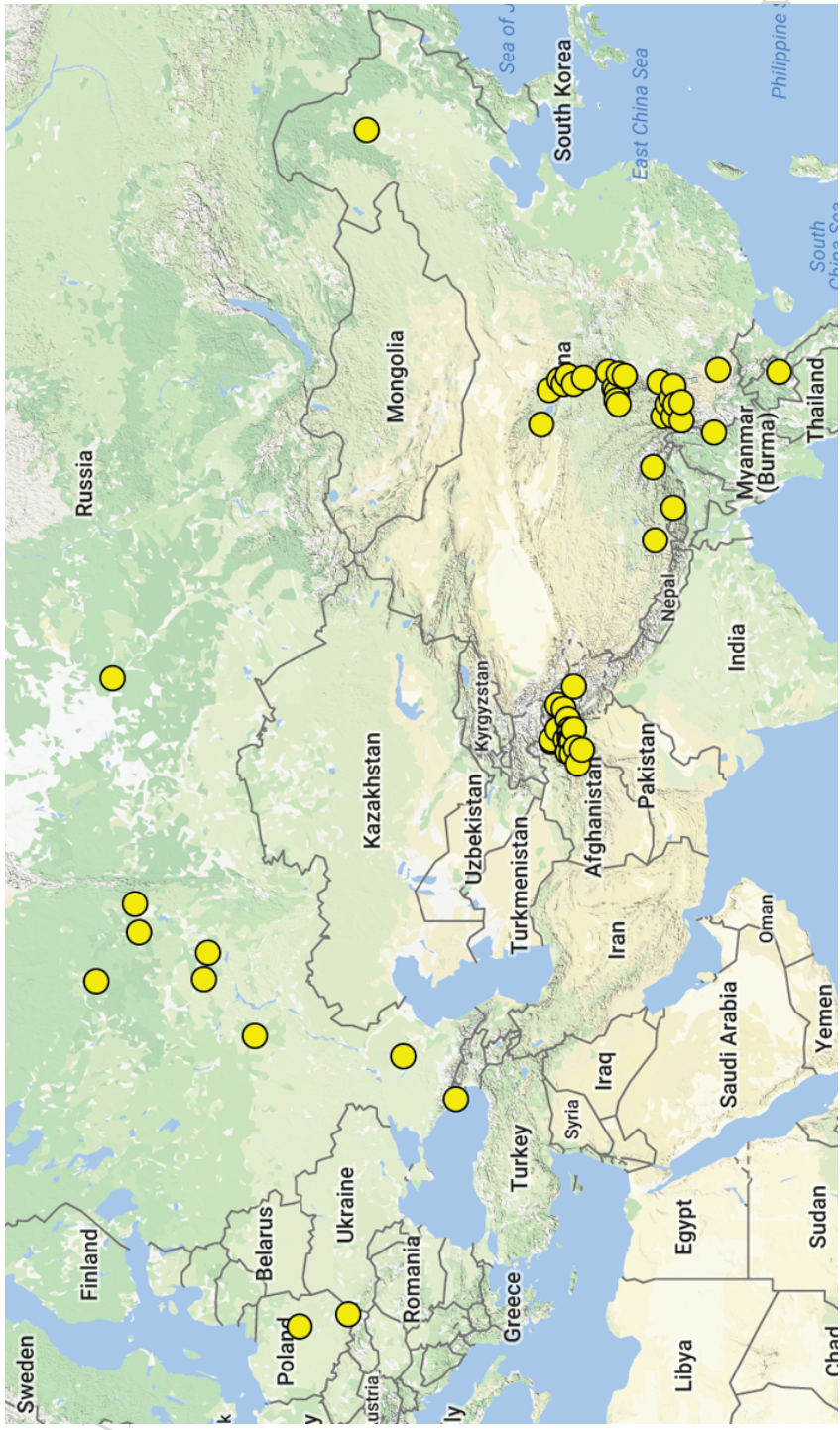

Figure 5. Distribution of retroflex affricates in Eurasia (Nikolaev et al. 2015)

Central European and Caucasian areas. They additionally make an appearance in several Uralic languages of the Russian Federation that have two series of affricates in the inherited lexicon (these are the only languages that most probably inherited retroflex affricates from their proto-language, cf. § 3.3.1). Their different densities in the West and in the East, however, are in line with our observations that the European and Caucasian affricate areas are very different from the eastern ones. In fact, the European area is probably not a true area at all, but rather an outcome of independent developments in Slavic languages (which are sketched in § 3.3.2), while the affricate richness of the Caucasus is not the result of borrowing of segments or of specific contact-induced sound changes. Rather, it appears to be the outcome of a more general process of saturation of consonant inventories (which may or may not be an areal feature in its own right) as analysed by Lindblom & Maddieson (1988). Therefore, in the subsequent discussion we will concentrate on the Hindukush and Eastern Himalayan areas.[10]

3. Sound-change processes giving rise to affricate-rich areas

3.1 The Hindukush area

Even though the Hindukush area is a *locus classicus* for affricate-rich linguistic areas, its sound-change history is rather poorly understood.

The situation is most dire with Burushaski, a language isolate or a group of closely-related languages, since in the absence of comparative evidence we cannot reconstruct its proto-history and cannot be sure even whether it inherited its numerous affricates from its proto-language or, as most other languages of the region, acquired them due to processes of sound change.

The early history of Nuristani, Iranian, and Dardic is much better known. There is no doubt that the denti-alveolar and postalveolar affricates in these languages arose due to (i) historical development of the Proto-Indo-European segments that are traditionally reconstructed as 'palatal velars' and (ii) processes of palatalisation of velar and coronal stops. A full list of the relevant sound changes and of similar sound changes in other Indo-European languages is given in Kümmel (2007). What is left to show is how the languages in question acquired the third, viz., retroflex,

10. With this proviso, it is still remarkable to which extent affricate-rich languages in Eurasia tend to come in rather tight areal clusters. This is apparently also the case in South America, which shows the areal effects in the distribution of affricate-rich inventories, based on data from the South American Phonological Inventory Database (Michael et al. 2015). See Figure 8 in Appendix D.

series of affricates. The relevant collection in Kümmel (2007) is evidently incomplete,[11] which reflects a real gap in the literature.

The only existing comparative grammar of East Iranian languages (Edelman 1986) does not provide a clear overview of the process of emergence of retroflex affricates in the Iranian languages of Hindukush, although it does suggest that they may have developed under the influence of Burushaski.

There is no general treatment of Dardic and Nuristani historical phonology, but studies of the phonologies of individual languages of the region containing diachronic analyses show that retroflex affricates have diverse origins. The possible sources, in addition to those listed by Kümmel (2007, cf. fn. 3), include *pr and *k(ʷ)s in Dardic (Konow 1936; Kogan 2015), *kr, *kl, and *ks in Nuristani (Hegedűs & Blažek 2010; Hegedűs 2012), and affricates at other places of articulation that underwent retroflexivisation (Hegedűs 2012). Arsenault (2012) also argues that some retroflex affricates in Dardic may have arisen due to the process of retroflex harmony.[12]

The phonological history of Balti, however, is straightforward in this respect. It evidently preserved dental and alveolo-palatal affricates from a western dialect of Old Tibetan and acquired some new ones by the process of palatalisation (Backstrom 1994). Several of the Western Tibetic varieties (Eastern and Western Balti, Purki, and Ladakhi) acquired retroflex stops due to the process of simplification of consonant clusters.[13] However, only one variety, Western Balti, also acquired retroflex affricates and consequently became affricate-rich. The process is described as follows by Backstrom (1994: 49, emphasis ours):

> Western Balti has recently begun undergoing a change similar, but not identical, to the more advanced one in Ladakhi. So far, the change has only affected clusters which began with proto velar stops, not the labials or dentals. Furthermore, instead of producing retroflexed stops, the change in Western Balti is producing retroflexed affricates. Whether this is only a stage toward stop production, or a more permanent arrangement, only time will tell. *Areal linguistic features, however, seem to lend strength to the idea that this development may indeed be different, not only in its stage of advancement, but also in kind, from that seen in Ladakhi. Baltistan, especially*

11. Only the following sound changes are listed ('ʈ' and 'ʂ' in Kümmel's notation correspond to /ʈ/ and /ʂ/): Iranian: *tʃ > tʂ /_ Wakhi; Nuristani: *ts > tʂ / _r>0 Common Nuristani; Dardic: ʂ > tʂh /#_w Khowar; ʂt > tʂ /_ Bashkarik.

12. I.e., sound change of the form tɕ, tʃ > ʈʂ / _Vʂ

13. Cf. the typical sound changes:

pʰ.r, kʰ.r > ʈʰ / C,#V
p.r, k.r > ʈ / C,#_V
b.r, d.r > ʈ / #_V (or ɖ / #_V in the Western and Amdo varieties without devoicing)
b.r > ɖ / C_V

> *western Baltistan, is partially surrounded and, to a considerable extent, infiltrated by speakers of Shina, a Dardic (Indo-Aryan) language which, like several other languages of northern Pakistan, have an abundance of retroflexed sibilants and affricates.* This areal feature could well have provided the pressure which is currently influencing the development of similar affricates in Western Balti.

The analysis of these sound changes shows that if denti-alveolar and postalveolar affricates can arise from simple stops that underwent palatalisation, retroflex affricates in the absolute majority of cases arise from consonant clusters (especially clusters with rhotics) or other affricates and are not the first affricates in their respective inventories. Therefore, it may be hypothesised that presence of consonant clusters and/or affricates in the inventory is a prerequisite for the emergence of retroflex affricates, which explains their strong link with affricate-rich inventories both in Eurasia (cf. the map in Figure 5) and South America (cf. the map and the list of affricate-rich inventories in Appendix D).

The areal clustering of retroflex affricates may be interpreted in the framework of the Perceptual Magnet framework advanced by Blevins (2017). This framework explains the appearance of segments previously absent from a given inventory as a result of sound-change processes guided towards a particular end by a 'magnet' – a particular segment present in the neighbouring language or languages, present in the pool of synchronic variation for production of particular sound sequences, and winning the competition due to its being already established in the phonetic or even phonological lexicon of the speakers. Due to the acoustic similarity between fricatives and affricates it does not really matter in this case whether retroflex affricates are endogenous to the region or were brought to life by the presence of retroflex fricatives in some early Indo-Aryan variety. However, the difference between Western Balti and other Western Tibetan varieties suggests that the presence of retroflex affricates as magnets clearly plays a role.

Unfortunately, there are no historical data on particular cases of contact-induced sound changes establishing themselves in speech of individual speakers and then propagating through the population. However, the Perceptual Magnet framework seems to be the best way to account for spatial clustering of similar sound changes and sound-system convergence in the absence of largescale lexical-borrowing events, which were not recorded for the regions under discussion.

3.2 The Eastern Himalayan area

This phonological history of the Eastern Himalayan languages is not completely understood, but much ground has been gained in the recent years. Ongoing work on **Proto-rGyalrongic** seems to indicate that denti-alveolar and alveolop alatal

affricates were inherited while retroflex affricates stem from clusters with rhotics (Guillaume Jacques, p.c.).

There is also no reconstruction of **Proto-Qiangic** (this is a problematic linguistic grouping overall, and as Evans [2004] puts it, '[for Qiangic] it has been (and continues to be) difficult or impossible to establish regular sound correspondences…'), but preliminary work on some languages from this genus (Yu 2009) points to similar conclusions: 'Retroflexes in Lizu come mostly from two sources: complex velar clusters [velar + l/r/j], and *sr clusters.' There is also one word exemplifying the change *bl > dz̧ / m V. Dental and postalveolar affricates partly arose due to palatalisation, but at least some of both kinds were inherited from the proto-language.

Proto-Lolo-Burmese (the proto-language of Achang, Ahi, Nuosu, and possibly Yongning Na)[14] is reconstructed with only denti-alveolar affricates (Bradley 1979). Matisoff (2003: 21) notes that retroflex affricates in contemporary languages "seem to be secondarily derived from proto-clusters with medial liquids", but does not give concrete examples. A change *rts(ʰ) > ʈʂ(ʰ) was proposed for Yongning Na by Jacques and Michaud (2011). It may be noted that Lolo-Burmese languages from outside the area of interest for the most part collapsed C+l and C+r clusters in one series or even merge them with C+j clusters without producing retroflex affricates (Bradley 1979; Matisoff 2003). The furthest ones – Rangoon and Myeik Burmese – have only one series of affricates. As will be shown in the next section, some Lolo-Burmese languages also show a tendency to lose retroflex affricates after acquiring them.

The affricate history of **Chinese** consists of three stages (Baxter 1992):

1. At the Old Chinese stage, the language is thought to have had only dentialveolar affricates.
2. Middle Chinese acquired (i) palatal affricates from the palatalisation of velar and dental stops and (ii) retroflex affricates (from clusters consisting of coronals and rhotics).
3. The Middle Chinese affricate inventory was simplified in different ways, as seen in different contemporary varieties; these processes will be described in more detail in the next section.

Therefore, it seems that Middle Chinese (the sources for which are dated to the late 6th-early 7th c.) is the earliest documented case of the presence of retroflex affricates – or any retroflex segments at all – in the region. Based on lexical and grammatical similarities, it is often assumed that as an early stage Chinese was in contact with 'Altaic' or 'Transeurasian' languages in the north-west and with

14. Which some scholars classify as Qiangic (Jacques & Michaud 2011).

Hmong-Mien and possibly Mon-Khmer languages in the south (Starostin 2008; Niederer 1998; LaPolla 2001). On Turkic, Mongolic, and Hmong-Mien see below; Proto-Tungusic (Tsintsius 1949),[15] and Proto-Mon-Khmer (Shorto et al. 2006) are reconstructed without retroflex segments.

The phonological history of the Eastern Himalayan **Tibetic** varieties is relatively well known due to our knowledge of their proto-language (Hill 2010) and intensive fieldwork conducted in the region in the last two decades (Hongladarom 1996; Makley et al. 1999; Sun 2003; Haller 2004; Huber 2005; Sun 2006; Bartee 2007; Chirkova 2008; Suzuki & mTshomo 2009; Suzuki 2009; Chirkova 2010; Jacques 2011). Two main types of sound change that account for the emergence of retroflex affricates are recorded in the literature:

1. Retroflex affricates arising from clusters with rhotics:
   - k(ʰ)r > ʈʂ(ʰ) (Themchen, rGyalthang, Kami)
   - gr > ʈʂ (Themchen, Kami, Zhongu)
   - gr > ɖʐ[16] (Themchen, Kami, Zhongu)
   - p(ʰ)r > ʈʂ(ʰ) (Themchen, Kami, Zhongu)
   - dr > ʈʂ (Themchen, Kami, Zhongu)
   - dr > ɖʐ (Themchen, Kami, Zhongu)
2. Retroflex affricates arising from alveolo-palatal affricates before non-front vowels:
   - tɕ(ʰ) > ʈʂ(ʰ) (rGyalthang, Melung)
   - dɕ > ɖʐ (rGyalthang, Melung)

Thus, the Eastern Himalayan Tibetic varieties developed retroflex affricates from clusters with rhotics or by splitting the alveolo-palatal affricate series. The difference between eastern and more western Tibetic varieties is therefore the same as between Western Balti and more eastern varieties: while the central Tibetic dialects mostly changed the C+r clusters into retroflex stops, those peripheral varieties that did not retain them changed them into retroflex affricates, presumably under areal pressure. Melung Tibetan presents an interesting case since it aligns with central Tibetic dialects in converting C+r clusters into retroflex stops but nevertheless acquired retroflex affricates through conditioned shift of alveolo-palatal affricates (cf. the discussion of the Mongolic data below).

15. Some Tungusic languages of China, such as Kilen (Zhang 2013) acquired a series of retroflex affricates under the influence of Mandarin Chinese.

16. In most Tibetic varieties, the reflexes of voiced obstruents in syllable initials are split according to whether they were also word-initial (in which case they usually became voiceless with the concomitant development of a low tone on the following vowel) or preceded by another consonant.

It is important to point out that, to our knowledge, sound changes transforming stop-plus-rhotic clusters into retroflex affricates (or any other affricates) have not been recorded outside of the Hindukush and the Eastern Himalayan areas.

Earlier reconstructions of **Proto-Hmong-Mien** (Niederer 1998) included retroflex affricates in the inventory of proto-initials. However, in a more recent analysis (Ratliff 2010), they are also derived from clusters with rhotics and affricates at other places of articulation.

Discussing the phonological history of Qinghai-Gansu **Mongolic** languages, Nugteren (2011: 218) notes that "under the influence of Chinese and/or Tibetan, many QG languages have by now developed two or three sets of affricates". The processes leading to this were diverse. In Mongghul, the proto-Mongolic affricates *č and *ǰ (characterised as 'palatal' by Janhunen [2003]) mostly became /tʃʰ/ and /tʃ/ with retroflex affricates coming from borrowings. In Dongxian, *č and *ǰ became /ʈʂʰ/ and /ʈʂ/, while /tʃʰ/ and /tʃ/ developed due to palatalisation of *t and *d. In Mangghuer, *č and *ǰ split into /tʃʰ, tʃ/ and /ʈʂʰ, ʈʂ/ partly conditioned by the subsequent vowels, with front vowels favouring postalveolar affricates. Finally, the QG languages developed alveolar affricates by way of strengthening of *s in initial and intervocalic positions, as well as acquiring them from borrowings.

The presence of retroflex affricates in the Turkic language Salar is explained by Dwyer (2007: 267–268) as a result of loanword incorporation:

> The retroflex series, now phonemic to Salar, was borrowed from Northwestern Chinese (also with Amdo Tibetan influence) within the last century. … Northwestern Chinese and Amdo Tibetan both distinguish three series of spirants: apical/alveolar, palatal, and retroflex. Salar already has a partial alveolar series (/s, z/ but not /ts, dz, ʃ, tʃ/) and a full palatal series; since at least 15–25% of Salar's vocabulary consists of Chinese and Tibetan loans, it is unsurprising that the retroflex series has also become phonemic.[17]

Data from Salar and Qinghai-Gansu Mongolic languages show that, unlike languages from other families in the Eastern Himalayan affricate-rich zone, they developed retroflex affricates exclusively from other affricates and/or acquired them from lexical borrowings (which is logical given the aversion of Altaic/Transeurasian languages to initial consonant clusters [Robbeets 2017]).

17. It may be seen from this remark that the variety described by Dwyer differs from that described by Tenishev (1976), which included a full alveolar series, including affricates, and was used for our sample.

3.3 Retroflex affricates outside the Hindukush and the Eastern Himalayan area: Uralic and Slavic languages

In § 2.6, we discussed the fact that the distribution of affricate-rich inventories in Eurasia is correlated with that of retroflex affricates (shown in Figure 5), and our discussion of the pathways to affricate richness would be incomplete if we did not take into account the areal cluster of languages with retroflex affricates outside of the affricate-rich areas. We look at this in § 3.3.1. It is also necessary to investigate the emergence of affricate richness in Slavic languages in more detail (§ 3.3.2).

3.3.1 *Uralic*

The big outlying cluster of languages with retroflex affricates comprises Finno-Ugric languages (Komi-Zyrian, Komi-Yodzyak, Komi-Permyak, Udmurt, Eastern Mari, Erzya), and Proto-Uralic has been reconstructed with a retroflex affricate *t͡ʂ (Janhunen 1982).[18] Some of these languages may have innovated an alveolo-palatal affricate series, and due to recent borrowings from Russian (whose affricate inventory consists of /t͡s̪/ and /t͡ɕ/) acquired the dental affricate /t͡s̪/. These borrowings may potentially lead to another affricate-rich area with a reversed origin history: (i) one original series of retroflex affricates, (ii) an additional series developed due to internal processes, and (iii) the areally-imposed dental affricate. It is also possible, however, that their affricate inventories will become even more Russianised and will consist of a denti-alveolar and an alveolo-palatal series, as happened with Standard Moksha and Hill Mari.

3.3.2 *Slavic*

Similarly to the Indo-Iranian languages, the Slavic languages entered the historical era with denti-alveolar and postalveolar affricates resulting from what is traditionally called in the literature the first, second, and third palatalisations of velars (Carlton 1991). The subsequent history of individual languages included numerous additional palatalisation processes, which for the most part did not result in the emergence of additional places of articulation. However, there are several exceptions, of which Polish and Serbo-Croatian are representative examples.

Polish innovated an additional series of alveolo-palatal affricates /t͡ɕ/ and /d͡ʑ/ as a result of palatalisation of /t/ and /d/. By comparison, in contemporary Russian palatalised /tʲ/ and /dʲ/ either retained their additional articulation or, in the Moscow

18. We thank Ante Aikio for pointing out to us alternative views on the reconstruction of Proto-Uralic alveolo-palatal affricates. In particular, traditional reconstructions contain an alveolo-palatal affricate, which is disputed by Janhunen (1988). More recently, Zhivlov (2014) has analyzed the traditionally reconstructed alveolo-palatal sibilant as an affricate at the same place.

dialect, started shifting towards /ts/ and /dz/. After that, according to Hamann (2004), following the analysis proposed by Hall (1997) for early Indo-Aryan, the Polish affricate inventory consisting of /ts/, /tʃ/, /tɕ/, and their voiced counterparts became oversaturated due to the high acoustic similarity; consequently, postalveolars and alveolo-palatals and postalveolars unconditionally shifted to more 'grave sounding' retroflexes (the original analysis was proposed for fricative segments, but it naturally carries over to affricates). It must be noted that the analysis by Hall (1997) was predicated on the thesis that it is impossible to have both /ʃ, ɕ/ in an inventory,[19] which is patently wrong (EURPhon has 15 languages with this combination, including Nantong Chinese, Standard Tatar, Ter Saami, and Luxembourgish). However, a relaxed version of this hypothesis (the combination of palato-alveolars and alveolo-palatals is likely to be unstable) seems tenable, albeit in need of additional testing.[20]

The scenario in Serbo-Croatian is very similar except for the source of alveolo-palatals (they are supposed to stem from palatal stops /c/ and /ɟ/) and the fact that the shift of palato-alveolars to a more 'grave' pronunciation was achieved not by means of retroflexion, but rather of lip rounding. The resulting acoustic similarity of Serbo-Croatian labialised palato-alveolar affricates and retroflex affricates has led to a discussion in the literature and the competing analyses include /tʃ, dʒ/ vs. /tʃʷ, dʒʷ/, /tʃʷ, dʒʷ/ vs. /tɕ, dʑ/, and /ʈʂ, ɖʐ/ vs. /tɕ, dʑ/ (Petrović & Gudurić 2010; Morén 2006).

These data corroborate the thesis that retroflex affricates are somewhat 'marked' in that they are usually latecomers to affricate inventories. They also show that it is possible for languages to acquire retroflex fricative and affricate series without any kind of areal pressure. As Lower Sorbian and Vlax Romani demonstrate, however, it is equally possible to remain at the stage of a superdense palatal region (both these languages contrast /tʃ, dʒ/ with /tɕ, dʑ/), and consequently, it seems that retroflex affricates are much more likely to emerge in a contact scenario.

19. "One generalization that I assume to be exceptionless is summarized in (3):

(3) No language can contrast palatoalveolars and alveolopalatals." (Hall 1997: 205)

20. Cf. the theoretical discussion of the optimal structure of fricative inventories in (Boersma & Hamann 2007).

4. Languages outside the Hindukush and the Eastern Himalayan linguistic areas tend to lose retroflex affricates

The data on the spontaneous emergence of retroflex affricates in Slavic (and presumably in Uralic) suggests that the observed areal patterning has two aspects. The first is positive, so to speak: retroflex affricates tend to spread horizontally within inventories that already have affricates. The second is negative: retroflex affricates are predicted to be unstable without areal support. Data for the total disappearance of certain types of segments is understandably scarce, as there is usually no evidence to suggest they were there in the first place, but what data we do have suggests that, unlike coronals – which are easily acquired and preserved in sound inventories, and which are relatively hard to lose (Blevins 2009) – retroflex affricates are both hard to acquire and hard to preserve. More precisely, retroflex affricates tend to merge with affricates at other places of articulation.[21]

4.1 Croatian

In some varieties of Serbo-Croatian spoken in Croatia and Bosnia, the tripartite system /ts, ʈʂ, tɕ/ was simplified into a bipartite one: /ts, tɕ/. A phonetic analysis of this merger and a theoretical interpretation in the framework of Functional Optimality Theory was presented by Čavar (2011).

4.2 Uralic

Although the retroflex affricate *ʈʂ (Janhunen 1982) is confidently reconstructed for Proto-Uralic, it was lost in the Finnic, Hungarian, and Saamic branches and several languages of the Samoyedic branch, that is, nearly everywhere except for a small region west of the Urals, where its preservation may be due to mutual

21. We have not systematically studied deaffrication processes in the languages of Eurasia, but there is evidence for processes turning affricates into stops (as in Japonic and in the Southern Tungusic language Uilta spoken on Sakhalin island) or fricatives (as in Burmese and Yanghao Hmong). Numerous examples are collected in Kümmel (2007: 67–75). Absent a systematic study of deaffrication, we take these examples as possible and very tentative evidence of the relative instability of affricates in general, although given further data and analysis, it may turn out that such deaffrication processes themselves show areal effects, as perhaps suggested by the relative proximity of, e.g., Japonic and Uilta, on the one hand, and Yanghao Hmong, Burmese, and Yunnan Sinitic varieties, on the other (see § 4.4 below). We thank Guillaume Jacques and José Andrés Alfonso de la Fuente for discussing these issues with us.

areal support between languages of three different branches of the family (Ugric, Mordvinic, Permic, and Mari).

4.3 Lisu

The most conservative Central Lisu varieties have retroflex affricates in three places of articulation (a plain voiceless series /ts, ʈʂ, tɕ/ is accompanied by a voiceless aspirated and a voiced one). In Southern Lisu, they merged with alveolar fricatives and affricates, while in Northern Lisu they became allophones of alveolo-palatal fricatives and affricates (Bradley 2017).

4.4 Chinese

The most extensive data on the loss of retroflex affricates is provided by varieties of Chinese, both at the level of large dialect groups (some of which lost retroflex affricates completely) and at the level of the varieties inside the rest of the groups. We will give an overview of the distribution of retroflex affricates among the major dialect groups, and then turn to a more detailed description of several varieties of Mandarin Chinese. The process of retroflex-affricate loss in Chinese was in many cases a part of a more general process of loss of all retroflex segments, but at the moment there is not enough data to estimate the relationship between these two processes.

4.4.1 *Loss of retroflex affricates in non-Mandarin varieties*

Retroflex affricates were completely lost in Wu (Simmons 1999), Min[22] (Norman 1974), Hakka (except for the dialect of Sandu), and Yue (Yan 2006), and were lost in some varieties of Xiang, Gan, and Ping (Yan 2006). The map of the distribution of major Chinese dialect groups reproduced in Figure 6[23] shows that retroflex affricates progressively disappear along the north-west – southeast axis, that is, as we leave the area close to the Eastern Himalayas and approach the South China Sea.

22. The Min dialects are considered to contain a pre-Middle-Chinese lexical layer, which complicates their phonological history (Norman 1991).

23. The status of Jin as a major dialect group is disputed (Kurpaska 2010), and it patterns with Mandarin in its inventory of initials.

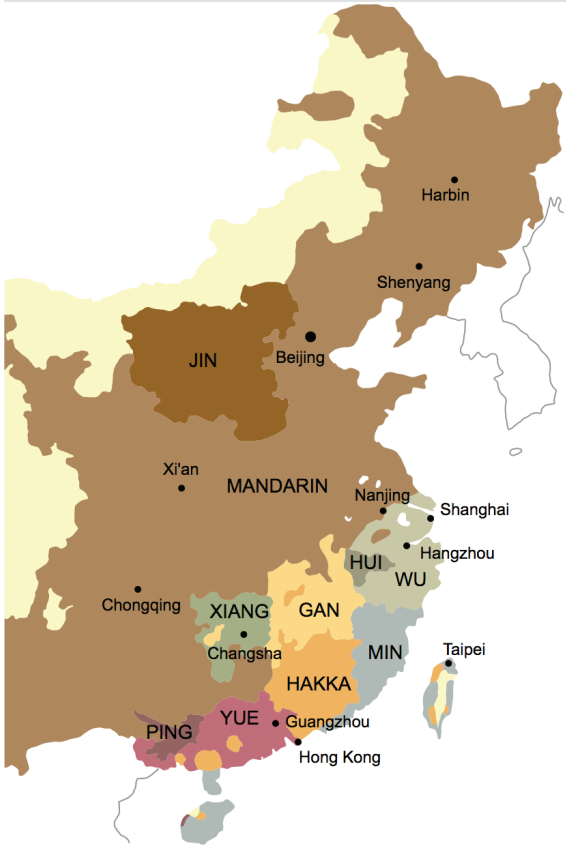

Figure 6. Geographical distribution of major Chinese dialect groups (Wikimedia Commons)

4.4.2 *Loss of retroflex affricates in the varieties of Mandarin Chinese*

Proto-Mandarin, the ancestor language of the contemporary Mandarin varieties, is considered to have retained Middle Chinese retroflex affricates (Baxter 1999, 2000). The Beijing variety of Mandarin Chinese, the most widely known dialect from this group, which became the basis for the standard language, not only retained Middle Chinese retroflex affricates, but also saw some of the previously palato-alveolar affricates and retroflex stops become retroflex affricates[24] as well

24. It is often remarked in the literature that Beijing Mandarin retroflexes are not 'real' retroflexes, but rather 'apical post-alveolar' sounds (Lee & Zee 2003). However, this characterisation falls under the broad definition of retroflex segments as 'bunched' non-laminal postalveolars, which is actually used in practice to describe consonant inventories of languages outside South India where 'real' tongue-curled retroflexes are abundant (Hamann 2003; Ladefoged & Maddieson 1996).

(Chen 1976). However, many other, especially southern, varieties of Mandarin are known to have merged retroflex affricates with alveolar affricates (Norman 1988). Two representative case-studies are the Mandarin varieties of the southern province of Yunnan and the partial or complete loss of retroflex affricates in Xi'an, Hefei, and Yangzhou as analysed by Baxter (Baxter 1999).

4.4.2.1 *Retroflex initials in the Mandarin varieties of Yunnan*

Gui (2001) analysed the geographical distribution of retroflex affricates in the Yunnan varieties of Southwest Mandarin Chinese and traced their history in the dialect of Kunming, the provincial capital. Diachronically, it is interesting to note that the transition between the so-called 'Old Kunming Chinese' and 'Contemporary Kunming Chinese' that, among other things, involved merger of retroflex affricates with the alveolar ones, took place relatively recently – in the second half of the 20th century.

The geographical distribution of retroflex-less varieties in Yunnan shows a clear centre-vs-periphery pattern. The map in Gui (2001: 42) shows that the varieties that retained retroflex affricates are concentrated in the middle part of the province, while the varieties on the periphery, spoken in the regions bordering on Burma, Laos, Vietnam, and Guangxi and Guizhou provinces – and thus in contact with retroflex-less Sinitic and non-Sinitic languages – lost them.

The recent phonological shift in the speech of the provincial capital, which is situated in the eastern part of the central region, can be explained by the influx of immigrants from peripheral regions speaking non-Sinitic languages or more deviant varieties of Mandarin during the 20th century.

4.4.2.2 *The fate of retroflex affricates in Xi'an, Hefei, and Yangzhou*

In his paper on the reconstruction of retroflex affricates in Proto-Mandarin, Baxter (1999) reconstructs, among other processes, partial or complete loss of retroflex affricates in several Mandarin varieties.

1. Xi'an, a variety of Central Plains Mandarin, "was affected by… retroflex loss before *ə, *a, and the reflex of original *u" (Baxter 1999: 28).
2. Hefei, a variety of Lower Yangtze Mandarin (a dialect group situated further to the east and influenced by Wu dialects), was affected by the same processes of retroflex loss, but it had not previously underwent the change i > ə, which affected northern dialects, and therefore had fewer contexts for this process.
3. Finally, Yangzhou, another Lower Yangtze Mandarin variety situated even further to the east from Hefei, underwent retroflex loss in all environments.

It may be added that the dialect of Jinan, a Jilu Mandarin (previously known as Northern Mandarin) variety, situated not far from Beijing, retained retroflex affricates in all environments. Therefore, we again witness a north-west – southeast

cline (Jinan – Xi'an+Hefei – Yangzhou) along which dialects progressively lose affricates, although on a smaller scale (it must be noted that the centre of gravity for the retroflex initials in this case is not only the Eastern Himalayan region, but also the influential Beijing Mandarin area).

4.4.3 *Retroflex affricates in Chinese, an overview*

In order to have a more complete understanding of the processes of the developments of retroflex segments in Chinese dialects, we need more diachronic analyses, but the data presented above point to two cycles of retroflex loss in the history of Chinese:

1. Early Post-Middle-Chinese dialects groupings situated further to the southeast from the Eastern Himalayas and the centre of gravity of Middle Chinese in the Central Plains lost retroflex initials, either at the level of local proto-languages, or later as a result of an areally-spreading sound change.
2. After the secondary southward spread of the speakers of initially retroflex-rich Mandarin Chinese varieties, southern varieties of Mandarin, which came into contact with retroflex-less non-Sinitic languages and dialects of Chinese that had already lost retroflex initials by that time, also started losing them, a process that seems to be still ongoing.

5. Conclusion

The analysis of the distribution of languages with affricates at three or more places of articulation shows that there two general pathways that lead to their emergence. Languages can either (i) develop affricates at peripheral places of articulation: labiodental and/or velar and post-velar (Standard and Swiss German, Kartvelian languages); or (ii) develop affricate-dense inventories with three or more places of articulation in the coronal-palatal range. Three additional observations can be made:

1. In the vast majority of cases, languages follow only one of these pathways.
2. The pathway involving affricate-dense inventories is much more widespread.
3. Affricate-dense inventories are extremely prone to contain retroflex affricates, which makes the latter a very strong correlate of affricate richness in general.

The analysis of diachronic pathways leading to affricate-rich inventories show that the emergence of retroflex affricates is in the majority of cases the last step towards affricate denseness: retroflex affricates are diachronically 'marked' in that they usually appear in inventories that already have affricates at two or more different places of articulation. Two scenarios were identified.

In the case of several Slavic languages and possibly some early Indo-Aryan varieties, tripartite fricative and affricate systems with a super-dense palatal region /s, ʃ, ɕ, ts, tʃ, tɕ/ saw the acoustically-based dissimilation of their middle elements, which led to the emergence of a more 'grave'-sounding retroflex fricative and affricate series /s, ʂ, ɕ, ts, tʂ, tɕ/.

In many other cases, however, the change was not spurred by any systemic factor and seems to have a purely areal basis. Numerous languages in the Hindukush and the Eastern Himalayan linguistic areas innovated a series of retroflex affricates starting from a 'maximally dispersed' system consisting of /ts/ and /tɕ/. The sound changes involved in this process (simplification of consonant clusters, especially those with rhotics, yielding retroflex affricates and retroflexivisation of /tɕ/) are not attested outside these regions.

Retroflex affricates, therefore, provide some of the most compelling evidence to date of areally-induced sound change. Compared to the relatively few cases of the internally-motivated appearance of retroflex affricates in Slavic languages, there is an impressive array of genera (and even phyla) whose languages underwent certain processes of sound-change that led to the emergence of retroflex affricates exclusively inside the Hindukush and the Eastern Himalayan regions.[25] Thus, our conclusions support the position of Edelman (1980), Tikkanen (1988, 1999, 2008), and Liljegren (2017), who argued that Hindukush is a *bona fide* linguistic area.

Finally, we were able to show, although only in a preliminary fashion, that languages outside of areas where retroflex affricates are common tend to lose them. This thesis is corroborated by the history of Croatian and Bosnian varieties of Serbo-Croatian, most branches of the Uralic family, several varieties of Lisu, and a large number of Chinese varieties situated to the south and south-east of the core retroflex areas in the Eastern Himalayas and North China Plain.

Thus, it is highly probable that affricate-rich areas not only facilitate the emergence of retroflex affricates in the nearby languages, but also help retain them. It is an interesting open question whether affricate richness or affricate density as abstract features of inventories can also be areally induced or strengthened. For instance, in the Eastern Himalayan area, there are several languages, such as Sangdam Tibetan, that are affricate-dense even though they do not have retroflex affricates. The processes leading to the emergence of such inventories are unclear. Even more problematic is the situation in the Caucasus where no less than three types of affricate richness are attested, and a general trend toward oversaturated

25. However, we still do not know the origins of retroflex affricates in Burushaski (inside the Hindukush area) and of the voiceless *ʈʂ reconstructed for Proto-Uralic, which is responsible for the single pocket of non-affricate-rich languages with retroflex affricates outside the Himalayas.

consonant inventories is evident (Grawunder 2017), but both facts at present lack any convincing areal interpretation.

More broadly, this paper has been a probe for distributional typology in the domain of phonology, asking 'what's where why when?'. The property of interest, affricate-rich and affricate-dense phonological inventories, has been shown to have different subtypes, to be geographically limited to a number of areas within the Eurasian macro-area, and to have developed through a limited number of pathways. This last in fact constitutes an area-specific diachronic typology of sound inventory complexification in a particular domain. Importantly, while the strong areal signal points to a contact-based explanation ('event-based', in Bickel's [2015] terminology), a finer-grained examination leads to the conclusion that these areas are not equally amenable to the same type of contact-based explanation. As a result, we conclude that distributional typologies can and should address the actual histories of languages and their speakers.

Of course, it remains to be seen to what extent these results can be generalized beyond Eurasia to macro-areas with affricate-rich and affricate-dense inventories (for instance, they also show a spatially clustered distribution in South America), a question we leave to specialists in these areas. In essence, however, such dense areal studies are necessary in order to provide an empirical basis for teasing apart the relative contributions of inheritance, language contact, and sound change to the observed diversity (and uniformity) of sound systems in the world's languages.

Acknowledgements

We are grateful to Ante Aikio, Nino Amiridze, Marko Crnobrnja, José Andrés Alfonso de la Fuente, Guillaume Jacques, Peter Piistanen, Juho Pystynen, Willem Vermeer, Rémy Videraz, Theo Yeh, and Mikhail Zhivlov for their comments on a draft of this paper. We also thank the audience at the 2017 (Zurich) meeting of the *Societas Linguistica Europaea*, the members of the Dynamics of Language Lab of the Hebrew University of Jerusalem, and the members of the Israel Phonology Circle, in particular Outi Bat-El, Evan-Gary Cohen, and Hadas Yeverachyahu.

References

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

Backstrom, Peter C. 1994. *A phonological reconstruction of Proto-Western Tibetan*. Arlington: University of Texas at Arlington MA thesis.

Bartee, Ellen Lynn. 2007. *A grammar of Dongwang Tibetan*. Santa Barbara: University of California, Santa Barbara dissertation.

Baxter, William Hubbard. 1992. *A handbook of Old Chinese phonology* (Trends in Linguistics 64). Berlin/New York: Mouton de Gruyter. https://doi.org/10.1515/9783110857085
Baxter, William Hubbard. 1999. Reconstructing Proto-'Mandarin' retroflex initials. In Richard VanNess Simmons (ed.), *Issues in Chinese dialect description and classification* (Journal of Chinese linguistics: Monograph series 15), 1–35. Berkeley, California: Project on Linguistic Analysis.
Baxter, William Hubbard. 2000. Did Proto-Mandarin exist? *Journal of Chinese Linguistics* 28(1). 100–115.
Bickel, Balthasar. 2007. Typology in the 21st century: Major current developments. *Linguistic Typology* 11(1). 239–251. https://doi.org/10.1515/LINGTY.2007.018
Bickel, Balthasar. 2015. Distributional typology. In Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of linguistic analysis*, 901–923. Oxford University Press.
Blevins, Juliette. 2009. Another universal bites the dust: Northwest Mekeo lacks coronal phonemes. *Oceanic Linguistics* 48(1). 264–273. https://doi.org/10.1353/ol.0.0033
Blevins, Juliette. 2017. Areal sound patterns: From perceptual magnets to stone soup. In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*, 88–121. Cambridge University Press. https://doi.org/10.1017/9781107279872.006
Boersma, Paul & Silke Hamann. 2007. The evolution of auditory contrast (unpublished manuscript). http://www.fon.hum.uva.nl/paul/papers/ EvolutionOfContrast.pdf.
Bradley, David. 1979. *Proto-Loloish* (Scandinavian Institute of Asian Studies monograph series no. 39). London: Curzon Press.
Bradley, David. 2017. Lisu. In R. P. E. Sybesma, Wolfgang Behr, Yueguo Gu, Zev J. Handel, Cheng-Teh James Huang & James Myers (eds.), *Encyclopedia of Chinese language and linguistics*, Leiden/Boston: Brill.
Campbell, Lyle. 2017. Why is it so hard to define a linguistic area? In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*, 19–39. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781107279872.003
Carlton, Terence R. 1991. *Introduction to the phonological history of the Slavic languages*. Columbus, Ohio: Slavica Publishers.
Ćavar, Malgorzata. 2011. Merger of the place contrast in the posterior sibilants in Croatian (unpublished manuscript).
Chen, Matthew Y. 1976. From Middle Chinese to Modern Peking. *Journal of Chinese Linguistics* 4(2/3). 113–277.
Chirikba, Viacheslav. 1999. The West Caucasian material in "The North Caucasian etymological dictionary" by S. L. Nikolayev and S. A. Starostin. In Helma van den Berg (ed.), *Studies in Caucasian linguistics*, 152–170. Leiden: CNWS Publications.
Chirkova, Ekaterina. 2014. Phonological profile of Kami, the Tibetan dialect of Muli (manuscript). In Jackson T. -S. Sun (ed.), *Special issue of Language and Linguistics on lesser known dialects of Tibetan*. In press.
Chirkova, Katia. 2008. On the position of Baima within Tibetan: A look from basic vocabulary. In Alexander Lubotsky & Frederik Kortlandt (eds.), *Evidence and counter-evidence: Essays in honour of Frederik Kortlandt. Volume 2: General linguistics*, 69–91. Amsterdam: Rodopi.
Denwood, Philip. 1999. *Tibetan* (London Oriental and African language library v. 3). Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/loall.3
Dwyer, Arienne M. 2007. *Salar: A study in Inner Asian language contact processes* (Turcologica 37). Wiesbaden: Harrassowitz.

Edelman, Dzhoi. 1980. K substratnomy naslediyu Tsentralnoaziatskogo yazykovogo soyuza [On the substrate heritage of the Central Asian linguistic area]. *Voprosy yazykoznaniya* 5. 21–33.
Edelman, Dzhoi. 1983. *The Dardic and Nuristani languages*. Moscow: Nauka.
Edelman, Dzhoi. 1986. *Sravnitelnaya grammatika vostochnoiranskikh yazykov. Fonologiya* [Comparative grammar of East Iranian languages. Phonology]. Moscow: Nauka.
Evans, Jonathan. 2004. The reconstruction of Proto-Qiang verb inflection. In Ying-chin Lin, Fang-min Hsu, Chun-chih Lee, Jackson T. -S. Sun, Hsiu-fang Yang & Dah-an Ho (eds.), *Studies on Sino-Tibetan languages: Papers in honor of professor Hwang-cherng Gong on his seventieth birthday*, 201–238. Taipei: Instute of Linguistics, Academia Sinica.
Gordon, Matthew & Ayla Applebaum. 2006. Phonetic structures of Turkish Kabardian. *Journal of the International Phonetic Association* 36(2). 159. https://doi.org/10.1017/S0025100306002532
Gordon, Matthew Kelly. 2016. *Phonological typology*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199669004.001.0001
Grawunder, Sven. 2017. The Caucasus. In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*, 356–395. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781107279872.014
Gui, Ming Chao. 2001. *Yunnanese and Kunming Chinese: a study of the language communities, the phonological systems, and the phonological developments* (LINCOM studies in Asian linguistics 28). München: LINCOM Europa.
Hall, T. Alan. 1997. The historical development of retroflex consonants in IndoAryan. *Lingua* 102(4). 203–221. https://doi.org/10.1016/S0024-3841(96)00050-2
Haller, Felix. 2004. Dialekt und Erzählungen von Themchen: sprachwissenschaftliche Beschreibung eines Nomadendialektes aus Nord-Amdo (Beiträge zur tibetischen Erzählforschung Bd. 14). Bonn: VGH Wissenschaftsverlag.
Hamann, Silke. 2004. Retroflex fricatives in Slavic languages. *Journal of the International Phonetic Association* 34(1). 53–67. https://doi.org/10.1017/S0025100304001604
Hamann, Silke Renate. 2003. *The phonetics and phonology of retroflexes* (LOT 75). Utrecht: LOT.
Hegedűs, Irén. 2012. The RUKI-rule in Nuristani. In Benedicte Nielsen Whitehead (ed.), *The sound of Indo-european: phonetics, phonemics and morphophonemics* (Copenhagen studies in Indo-European vol. 4), 145–168. Copenhagen: Museum Tusculanum, University of Copenhagen.
Hegedűs, Irén & Václav Blažek. 2010. On the position of Nuristani within IndoIranian. Paper presented at the conference Sound of Indo-European 2 (Opava, Oct 2010).
Hill, Nathan W. 2010. An overview of Old Tibetan synchronic phonology. *Transactions of the Philological Society* 108(2). 110–125. https://doi.org/10.1111/j.1467-968X.2010.01234.x
Hongladarom, Krisadawan. 1996. Rgyalthang Tibetan of Yunnan: A preliminary report. *Linguistics of the Tibeto-Burman Area* 19(2). 69–92.
Huber, Brigitte. 2005. The Tibetan dialect of Lende (Kyirong): *A grammatical description with historical annotations* (Beiträge zur tibetischen Erzählforschung Bd. 15). Bonn: VGH Wissenschaftsverlag.
Jacques, Guillaume. 2014. A phonological profile of Cone. In Jackson T. -S. Sun (ed.), *Phonological profiles of little-studied Tibetic varieties* (Language and Linguistics monograph series 55), 269–330. Taipei: Institute of Linguistics, Academia Sinica.
Jacques, Guillaume & Alexis Michaud. 2011. Approaching the historical phonology of three highly eroded Sino-Tibetan languages: Naxi, Na and Laze. *Diachronica* 28(4). 468–498. https://doi.org/10.1075/dia.28.4.02jac

Janhunen, Juha. 1982. On the structure of Proto-Uralic. *Finnisch-ugrische Forschungen* 44. 23–42.

Janhunen, Juha. 2003. Proto-Mongolic. In Juha Janhunen (ed.), *The Mongolic languages* (Routledge language family series 5), 1–29. London: Routledge.

Kogan, Anton. 2015. Nekotorye voprosy geneticheskoj klassifikacii dardskih jazykov po dannym istoricheskoj fonetiki [Some issues of the genealogical classification of the Dardic languages based on historical phonology]. *Journal of Language Relationship* 13(12). 1–21.

Konow, Sten. 1936. Note on the ancient North-Western Prakrit. *Bulletin of the School of Oriental and African Studies* 8(2–3). 603–612. https://doi.org/10.1017/S0041977X00141230

Kümmel, Martin. 2007. *Konsonantenwandel: Bausteine zu einer Typologie des Lautwandels und ihre Konsequenzen für die vergleichende Rekonstruktion*. Wiesbaden: Reichert.

Kurpaska, Maria. 2010. *Chinese language(s): A look through the prism of the Great Dictionary of Modern Chinese Dialects* (Trends in linguistics: Studies and monographs 215). Berlin/New York: De Gruyter Mouton. https://doi.org/10.1515/9783110219159

Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford, UK/Cambridge, MA: Blackwell Publishers.

LaPolla, Randy J. 2001. The role of migration and language contact in the development of the Sino-Tibetan language family. In Alexandra Y. Aikhenvald & Robert M. W. Dixon (eds.), *Areal diffusion and genetic inheritance: problems in comparative linguistics*, 225–254. Oxford/New York: Oxford University Press.

Lee, Wai-Sum & Eric Zee. 2003. Standard Chinese (Beijing). *Journal of the International Phonetic Association* 33(1). 109–112. https://doi.org/10.1017/S0025100303001208.

Liljegren, Henrik. 2017. Profiling Indo-Aryan in the Hindukush-Karakoram: A preliminary study of micro-typological patterns. *Journal of South Asian Languages and Linguistics* 4(1). 107–156. https://doi.org/10.1515/jsall-2017-0004

Liljegren, Henrik & Erik Svärd. 2017. Bisyndetic contrast marking in the Hindukush: Additional evidence of a historical contact zone. *Journal of Language Contact* Online preprint. https://doi.org/10.1163/19552629-01002010

Lindblom, Björn & Ian Maddieson. 1988. Phonetic universals in consonant systems. In Victoria Fromkin, Larry M. Hyman & Charles N. Li (eds.), *Language, speech, and mind: studies in honour of Victoria A. Fromkin*, 62–78. London/New York: Routledge.

Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511753459

Maddieson, Ian & Kristin Precoda. 1992. *UPSID and PHONEME (version 1.1)*. Los Angeles: University of California at Los Angeles.

Makley, Charlene, Keith Dede, Hua Kan & Wang Qingshan. 1999. The Amdo dialect of Labrang. *Linguistics of the Tibeto-Burman Area* 22(1). 97–127.

Matisoff, James A. 2003. *Handbook of Proto-Tibeto-Burman: system and philosophy of Sino-Tibetan reconstruction* (University of California publications in linguistics v. 135). Berkeley: University of California Press.

Michael, Lev, Tammy Stark, Emily Clem & Will Chang (eds.). 2015. *South American phonological inventory database v1.1.4*. Berkeley: University of California. http://linguistics.berkeley.edu/~saphon/.

Moran, Steven, Daniel McCloy & Richard Wright (eds.). 2014. *Phoible online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://phoible.org/.

Morén, Bruce. 2006. Consonant-vowel interactions in Serbian: Features, representations and constraint interactions. *Lingua* 116(8). 1198–1244. https://doi.org/10.1016/j.lingua.2005.04.003

Nichols, Johanna. 1997. Modeling ancient population structures and movement in linguistics. *Annual review of anthropology* 26(1). 359–384. https://doi.org/10.1146/annurev.anthro.26.1.359

Niederer, Barbara. 1998. *Les langues Hmong-Mjen (Miáo-Yáo): phonologie historique* (LINCOM studies in Asian linguistics 07). München: LINCOM EUROPA.

Nikolaev, Dmitry. 2018. The Database of Eurasian Phonological Inventories: A research tool for distributional phonological typology. *Linguistics Vanguard* 4(1). https://doi.org/10.1515/lingvan-2017-0050

Nikolaev, Dmitry, Andrey Nikulin & Anton Kukhto (eds.). 2015. *The database of Eurasian phonological inventories (beta version)*. http://eurasianphonology.info/.

Norman, Jerry. 1974. The initials of Proto-Min. *Journal of Chinese Linguistics* 2(1). 27–36.

Norman, Jerry. 1988. *Chinese* (Cambridge language surveys). Cambridge, UK/New York: Cambridge University Press.

Norman, Jerry. 1991. The Min dialects in historical perspective. In William S. -Y. Wang (ed.), *Languages and dialects of China* (Journal of Chinese Linguistics Monograph Series 3), 323–358. Berkeley, California: Project on Linguistic Analysis.

Nugteren, Hans. 2011. *Mongolic phonology and the Qinghai-Gansu languages*. Utrecht: LOT.

Petrović, Dragoljub & Snežana Gudurić. 2010. *Fonologija srpskoga jezika* [Phonology of Serbian]. Beograd: Institut za srpski jezik SANU.

Ratliff, Martha Susan. 2010. *Hmong-Mien language history* (Studies in language change 8). Canberra: Pacific Linguistics.

Robbeets, Martine. 2017. The Transeurasian languages. In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*, 586–626. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781107279872.023

Shorto, Harry Leonard, Paul Sidwell, Doug Cooper & Christian Bauer. 2006. *A Mon-Khmer comparative dictionary* (Pacific linguistics 579). Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University.

Simmons, Richard VanNess. 1999. *Chinese dialect classification: A comparative approach to Harngjou, Old Jintarn, and common northern Wu*. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/cilt.188

Starostin, Sergei. 2008. Altaic loans in Old Chinese. In Alicia Sanchez-Mazas, Roger Blench, Malcolm D. Ross, Ilia Peiros & Marie Lin (eds.), *Past human migrations in East Asia: Matching archaeology, linguistics and genetics* (Routledge Studies in the Early History of Asia 5), 254–262. London/New York: Routledge.

Stolz, Thomas & Nataliya Levkovych. 2017. Convergence and divergence in the phonology of the languages of Europe. In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*, 122–160. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781107279872.007

Sun, Jackson T. -S. 2003. Phonological profile of Zhongu: A new Tibetan dialect of Northern Sichuan. *Language and linguistics* 4(4). 769–836.

Sun, Jackson T. -S. 2006. Special linguistic features of gSerpa Tibetan. *Linguistics of the Tibeto-Burman area* 29(1). 107–126.

Suzuki, Hiroyuki. 2009. Tibetan dialects spoken in Shar khog and Khod po khog. *East and West* 59(1/4). 273–283.

Suzuki, Hiroyuki & Tshering mTshomo. 2009. Preliminary analysis of the phonological history of Melung Tibetan. *Language and Linguistics* 10(3). 521–537.

Tenishev, Edgem R. 1976. *Stroj salarskogo jazyka* [Grammar of Salar]. Moscow: Nauka.

Tikkanen, Bertil. 1988. On Burushaski and other ancient substrata in northwestern South Asia. *Studia Orientalia* 64. 303–325.
Tikkanen, Bertil. 1999. Archaeological-linguistic correlations in the formation of retroflex typologies and correlating areal features in South Asia. In Roger Blench & Matthew Spriggs (eds.), *Archaeology and languages IV: Language change and cultural transformation*, 138–148. Abingdon, UK: Taylor & Francis. https://doi.org/10.4324/9780203208793_chapter_6
Tikkanen, Bertil. 2008. Some areal phonological isoglosses in the transit zone between South and Central Asia. In Israr-ud-Din (ed.), *Proceedings of the third international Hindu Kush cultural conference*, 250–262. Oxford/New York: Oxford University Press.
Tsintsius, Vera. 1949. *Sravnitelnaya grammatika tunguso-manchzhurskikh yazykov* [Comparative grammar of Tungus-Manchu languages]. Leningrad: Gosuchpedgiz.
Yan, Margaret Mian. 2006. Introduction to Chinese dialectology (LINCOM Studies in Asian linguistics 22). München: LINCOM Europa.
Yu, Dominic. 2009. *Lizu and ProtoTibeto-Burman*. Unpublished manuscript. University of Berkeley.
Zhang, Paiyu. 2013. *The Kilen language of Manchuria: Grammar of a Moribund Tungusic Language*: University of Hong Kong dissertation. https://doi.org/10.5353/th_b4985881
Zhivlov, Mikhail. 2014. Studies in Uralic vocalism III. *Journal of Language Relationship* 12. 113–148.

Appendices

A. Distribution of affricate-rich and non-affricate-rich languages in Eurasia

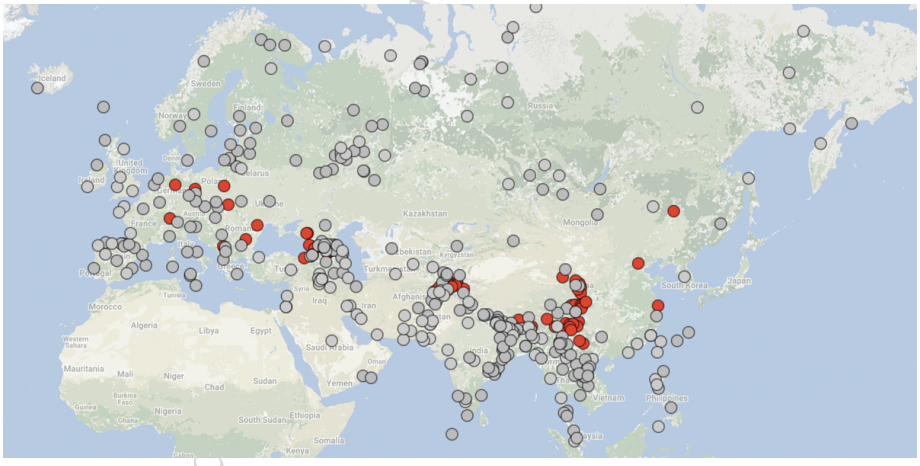

Figure 7. Affricate-rich (red) vs. non-affricate-rich (grey) languages in Eurasia (Nikolaev et al. 2015)

B. Affricate-rich languages of the Hindukush area

Phylum	Genus	Language name	Affricate inventory
Indo-European	Indo-Iranic	Southwest Pashai	ts, tʃ, dʒ, tʃʰ, dʒʰ, ʦ, ʦʰ
		Kati	ts, tsʲ, tsʷ, dz, tʃ, tʃʷ, dʒ, ʦ, ʦʷ
		Munji	ts, dz, tʃ, dʒ, ʦ, ʣ
		Shumashti	ts, tʃ, ʦ, ʦʰ
		Wotapuri-Katarqalai	ts, tsʰ, dz, ʦ, ʦʰ, dʒ
		Gawarbati	ts, tsʰ, tʃ, dʒ, ʦ, ʦʰ
		Ishkashimi	ts, dz, tʃ, dʒ, ʦ
		Dameli	ts, tsʰ, tʃ, tʃʰ, ʦ, ʦʰ
		Northern Kalasha	ts, tsʰ, dz, dzʰ, tʃ, tʃʰ, dʒ, dʒʰ, ʦ, ʦʰ, ʣ
		Palula	ts, tɕ, tɕʰ, ʦʰ
		Wakhti	ts, dz, tʃ, dʒ, ʦ, ʣ
		Kalkoti	ts, tɕ, tɕʰ, dʑ, ʦ
		Khowar	ts, tsʰ, dz, tɕ, tɕʰ, dʑ, ʦ, ʦʰ, ʣ
		Kalam Kohistani	ts, tsʰ, tʃ, tʃʰ, dʒ, ʦ, ʦʰ
		Ushojo	ts, dz, tɕ, dʑ, ʦ, ʣ
		Indus Kohistani	ts, tsʰ, tʃ, tʃʰ, dʒ, ʦ
		Shina	ts, tsʰ, tʃ, tʃʰ, dʒ, ʦ, ʦʰ
Burushaski	–	Werchikwar, Hunza, Nagar	ts, tsʰ, tɕ, tɕʰ, dʑ, ʦ, ʦʰ, ʣ
Sino-Tibetan	Tibetic	Balti (Kharkoo)	ts, tsʰ, tʃ, tʃʰ, dʒ, ʦ, ʦʰ, ʣ

C. Affricate-rich languages of the Eastern Himalayan area

Phylum	Genus	Language name	Affricate inventory
Sino-Tibetan	Tibetic	Shigatse Tibetan	ts, tsʰ, ʦ, ʦʰ, tɕ, tɕʰ
		Southern Cuona	ts, tsʰ, dz, ʦ, ʦʰ, ʣ, tɕ, tɕʰ, dʑ
		Northern Cuona	ts, tsʰ, dz, ʦ, ʦʰ, ʣ, tɕ, tɕʰ, dʑ
		Sangdam Tibetan	ts, tsʰ, dz, tɕ, tɕʰ, dʑ, cʑ, cʑʰ, ɟʝ
		Themchen Tibetan	ts, tsʰ, dz, ʦ, ʦʰ, ʣ, tɕ, tɕʰ, dʑ
		Melung Tibetan	ts, tsʰ, dz, tɕ, tɕʰ, dʑ, ʦ, ʦʰ, ʣ
		Dongwang Tibetan	ts, tsʰ, dz, ⁿdz, tɕ, tɕʰ, dʑ, ⁿdʑ, ʦ, ʦʰ, ʣ, ⁿʣ
		Rgyalthang Tibetan	ts, tsʰ, dz, ⁿdz, ʦ, ʦʰ, ʣ, ⁿʣ, tɕ, tɕʰ, dʑ, ⁿdʑ
		gSerpa	ts, tsʰ, ⁿdz, ʦ, ʦʰ, ʣ, ⁿʣ, tʃ, tʃʰ, dʒ, ⁿdʒ
		Khalong Tibetan	ts, tsʰ, ⁿtsʰ, ⁿdz, tʃ, tʃʰ, ⁿtʃʰ, dʒ, ⁿdʒ, ʦ, ʦʰ, ʣ, ⁿʣ

Phylum	Genus	Language name	Affricate inventory
		Kami Tibetan	ts, tsʰ, dz, ⁿdz, tʃ, tʃʰ, dʒ, ⁿdʒ, tɕ, tɕʰ, dʑ, ⁿdʑ, ʈʂ, ʈʂʰ, ɖʐ, ⁿɖʐ
		Labrang Tibetan	ts, tsʰ, dz, ⁿdz, ʈʂ, ʈʂʰ, ɖʐ, ⁿɖʐ, tɕ, tɕʰ, dʑ, ⁿdʑ
		Nyinpa Cone	ts, tsʰ, dz, ⁿdz, tɕ, tɕʰ, dʑ, ⁿdʑ, ʈʂ, ʈʂʰ, ɖʐ, ⁿɖʐ
		Zhongu Tibetan	ts, ʰts, tsʰ, ⁿtsʰ, dz, ⁿdz, ʈʂ, ʈʂʰ, ⁿʈʂʰ, ɖʐ, ⁿɖʐ, tʃ, ʰtʃ, tʃʰ, ⁿtʃʰ, dʒ, ⁿdʒ
		Baima	ts, tsʰ, dz, tʃ, tʃʰ, dʒ, ⁿdʒ, tɕ, tɕʰ, ⁿdʑ
	Sinitic	Kunming Chinese (SW Mandarin)	tɕ, tɕʰ, ʈʂ, ʈʂʰ, ts, tsʰ
		Standard Chinese (Beijing)	ts, tsʰ, tʃ, tʃʰ, tɕ, tɕʰ
		Nantong Chinese	ts, tsʰ, tʃ, tʃʰ, tɕ, tɕʰ
	Qiangic	Southern Pumi	ts, tsʰ, dz, tʃ, tʃʰ, dʒ, ʈʂ, ʈʂʰ, ɖʐ
		Lizu	ts, tsʰ, dz, tʃ, tʃʰ, dʒ, tɕ, tɕʰ, dʑ, ⁿtsʰ, ⁿdz, ⁿtʃʰ, ⁿdʒ, ⁿtɕʰ, ⁿdʑ
		Ersu	t̪s̪, t̪s̪ʰ, d̪z̪, ⁿt̪s̪ʰ, ⁿd̪z̪, ts, tsʰ, dz, ⁿtsʰ, ⁿdz, tɕ, tɕ, dʑ, ⁿtɕʰ, ⁿdʑ, ʈʐ, ʈʐʰ, ɖʐ, ⁿʈʐʰ, ⁿɖʐʰ
		Northern Qiang (Hongyan)	ts, tsʰ, dz, ʈʂ, ʈʂʰ, dʐ, tɕ, tɕʰ, dʑ
		Northern Qiang (Yadu)	ts, tsʰ, dz, ʈʂ, ɖʐ, tɕ, tɕʰ, dʑ
		Puxi	ts, tsʰ, dz, ʈʂ, ʈʂʰ, ɖʐ, tɕ, tɕʰ, dʑ
	rGyalrongic	Caodeng rGyalrong	ts, tsʰ, dz, ⁿdz, tʃ, tʃʰ, dʒ, ⁿdʒ, ʈʂ, ʈʂʰ, ɖʐ, ⁿɖʐ
		Japhug	ts, tsʰ, dz, ⁿdz, tɕ, tɕʰ, dʑ, ⁿdʑ, ʈʂ, ʈʂʰ, ɖʐ, ⁿɖʐ
	Lolo-Burmese	Achang (Lianghe)	ts, tsʰ, tɕ, tɕʰ, tʂ, tʂʰ
		Nuosu (Black Yi)	ts, tsʰ, dz, ⁿdz, ʈʂ, ʈʂʰ, ɖʐ, ⁿɖʐ, tʃ, tʃʰ, dʒ, ⁿdʒ, tɕ, tɕʰ, dʑ, ⁿdʑ
		Ahi	ts, tsʰ, dz, tɕ, tɕʰ, dʑ, ʈʂ, ʈʂʰ, ɖʐ
	Naxi	Yongning Na (Mosuo)	ts, tsʰ, dz, tɕ, tɕʰ, dʑ, ʈʂ, ʈʂʰ, ɖʐ
	–	Shixing	t̪s̪, t̪s̪ʰ, d̪z̪, ʈʂ, ʈʂʰ, ɖʐ, tɕ, tɕʰ, ⁿd̪z̪, ⁿɖʐ, dʑ, ⁿdʑ
Turkic	Oghuz	Salar	tsʰ, dz, tʃʰ, dʒ, tɕʰ, dʑ, ʈʂʰ, ɖʐ
Mongolic	Qinghai-Gansu	Mongghul	ts, tsʰ, tʃ, tʃʰ, ʈʂ, ʈʂʰ
		Mangghuer	ts, tsʰ, tʃ, tʃʰ, ʈʂ, ʈʂʰ
		Dongxiang	ts, tsʰ, tʃ, tʃʰ, ʈʂ, ʈʂʰ
	Daur	Daur	ts, tsʰ, tʃ, tʃʰ, ʈʂ, ʈʂʰ, tʃʷ, tʃʷʰ
Hmong-Mien	Hmongic	Hmong Njua	ts, tsʰ, ⁿts, ⁿtsʰ, tʃ, tʃʰ, ⁿtʃ, ⁿtʃʰ, ʈʂ, ʈʂʰ, ⁿʈʂ, ⁿʈʂʰ
		Xong	ts, tsʰ, ⁿts, ⁿtsʰ, ʈʂ, ʈʂʰ, ⁿʈʂ, ⁿʈʂʰ, tɕ, tɕʰ, ⁿtɕ, ⁿtɕʰ

D. Affricate-rich languages in South America

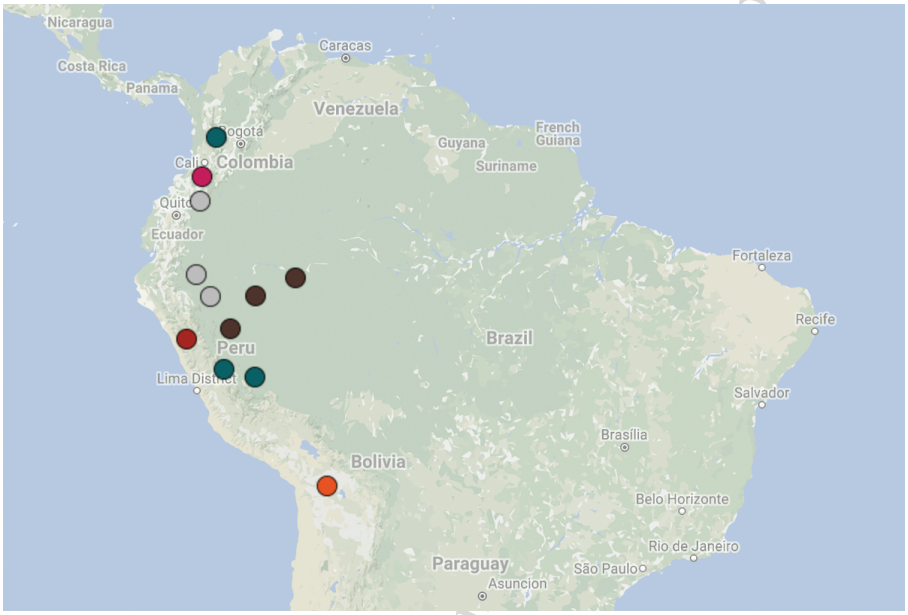

Figure 8. Distribution of affricate-rich languages in South America (Michael et al. 2015)

Table 9. Affricate-rich languages in South America

Phylum	Language	Affricate inventory
Panoan	Matís	ts, tʃ, ʈʂ
	Matsés	ts, tʃ, ʈʂ
	Shipibo	ts, tʃ, ɖʐ
Arawak	Chamicuro	ts, tʃ, ʈʂ
	Yánesha	ts, tʃ, ʈʂ
	Yine	ts, tʃ, cç
Quechua	Ancash Quechua	ts, tʃ, ʈʂ
Barbacoan	Guambiano	ts, tʃ, ʈʂ
Uru-Chipaya	Chipaya	ts, tsʾ, tsʰ, tʃ, tʃʾ, tʃʰ, ʈʂ, ʈʂʾ, ʈʂʰ
Isolates	Candoshi-Shapra	ts, tʃ, ʈʂ
	Muniche	ts, tʃ, ʈʂ
	Camsá	ts, tʃ, ʈʂ

E. Overview of sound changes producing retroflex affricates

Table 10. Sound change processes leading to retroflex affricates

Type	Source	Context	Language
Unconditioned shift	dʒ, tʃ	_	Polish; SerboCroatian; Waxi (Iranian)
	dʑ, tɕ	_	Dongxian (Qinghai-Gansu, Mongolic); ProtoMandarin
Shift under the influence of a rhotic segment	ts	r_	Yongning Na (Qiangic, SinoTibetan)
	ts	_r	Common Nuristani
Shift under the influence of a nonfront vowel	tɕ	_V[−front]	Tibetic
Retroflex harmony	tɕ, tʃ	_C[+retroflex]	Dardic
Simplification of velar+sibilant clusters	ks, kʲs, kʷs		Common Nuristani, Common Dardic
Simplification of clusters with rhotics	tr	#_V	Proto-MiddleChinese, ProtorGyalrongic, Hmong-Mien
	br, p(ʰ)r	_	Common Dardic, Tibetic, HmongMien
	gr, k(ʰ)r	_	Tibetic, Nuristani
Simplification of clusters with laterals	kl	_	Common Nuristani
Affricativisation of retroflex stops	ʈ	#_V	Proto-Mandarin
Affricativisation of retroflex fricatives	ʂ	_w	Khowar (Dardic)
Metathesis	ʂʈ	_	Bashkarik (Dardic)

Author queries

Please provide a complete reference for the citation '(Chirkova 2010), (Jacques 2011)' in this article.
Please provide a citation for the reference id "CIT0019 (Chirkova, Ekaterina. 2014), CIT0040 (Jacques, Guillaume. 2014)" since citation is missing in the article.

Authors' addresses

Dmitry Nikolaev (corresponding author)
Department of Linguistics
Hebrew University of Jerusalem
Mt. Scopus
Jerusalem, 9190501
Israel
dsnikolaev@gmail.com
dmitry.nikolaev@mail.huji.ac.il

Eitan Grossman