Kinship terminologies reveal ancient contact zone in the Hindu Kush

Abstract: The Hindu Kush, or the mountain region of northern Pakistan, northeastern Afghanistan and the northern-most part of the Indian-administered Kashmir region, is home to approximately 50 languages belonging to six different genera: Indo-Aryan, Iranian, Nuristani, Sino-Tibetan, Turkic and the isolate Burushaski. Areality research on this region is only in its early stages, and while its significance as a convergence area has been suggested by several scholars, only a few, primarily phonological and grammatical, features have been studied in a more systematic fashion. Cross-linguistic research in the realms of semantics and lexical organization has been given considerably less attention. However, preliminary findings indicate that features are geographically bundled with one another, across genera, in significant ways, displaying semantic areality on multiple levels throughout the region or in one or more of its sub-regions. The present study is an areal-typological investigation of kinship terms in the region, in which particular attention is paid to a few notable polysemy patterns and what appears to be a significant geographical clustering of these. Comparisons are made between the geographical distribution of such patterns and those of some other linguistic features as well as with relevant non-linguistic factors related to shared cultural values or identities and a long history of small-scale cross-community interaction in different parts of the region.

Keywords: diffusion; feature distribution; language contact; marriage patterns; polysemy; semantic typology; substratum

1 The Hindu Kush region

The Hindu Kush region, the Greater Hindu Kush, or the Hindukush-Karakorum as it has been referred to elsewhere (Heegård Petersen 2015: 23–24; Koptjevskaja-Tamm and Liljegren 2017: 215; Liljegren 2017), is here really a shorthand for the Pamir–Hindukush–Karakorum–Kohistan–Kashmir region (Bashir 2016: 264), an ethno-linguistically rich region, politically divided between Afghanistan, Pakistan and...
India. The geographically most salient feature of this region is its mountainous environment, especially vis-à-vis the Indo-Gangetic plains situated south of it. It is a transit zone of sorts between the cultural spheres of South Asia, Central Asia, West Asia and the Himalayas. The region as of today is largely characterized by its inclusion in the larger Islamic cultural and religious context, although in various forms – Sunni, Shiite and Ismailite and Nurbakhshi. There are only two notable exceptions to this near-complete Islamization. First, a few smaller Buddhist or partly-Buddhist groups reside in Ladakh, at the eastern fringe of the region, as defined here. Second, the tiny non-Muslim section of the Kalasha community is found in northern Pakistan’s Chitral valley, in the inner parts of the Hindu Kush (Heegård Petersen 2015: 13), all surrounded by exclusively Muslim populations.

Linguistically, the region is simultaneously the easternmost extension of Iranian languages, the northernmost extension of Indo-Aryan languages as well as the westernmost extension of Sino-Tibetan. Apart from those three phylogenetic components, the region is also home to Nuristani (a third branch of Indo-Iranian, along with Indo-Aryan and Iranian), at least two Turkic languages and the language isolate Burushaski. Thus, in all, six separate genera are represented, of which Indo-Aryan with its approximately 30 languages is the numerically overall dominant one and the geographically most widespread. Indo-Aryan in this particular region has been subject to a great deal of debate (Strand 2001: 251), concerning whether these languages are truly Indo-Aryan, should be considered a classificatory unit of their own, or represent (perhaps along with the Nuristani languages) a transitional group between Indo-Aryan and Iranian. A term often seen in reference to these languages is “Dardic”. However, hardly any modern linguists adhere to this term as anything else than a convenient umbrella term for a group of languages that are characterized by a few salient retentions from previous stages of Indo-Aryan (Morgenstierne 1974: 3), but that also happen to share some contact-related developments (Bashir 2003: 821–822). Contact in that case includes mutual contact between the various Indo-Aryan linguistic communities as well as significant contact with adjacent communities belonging to other genera (Liljegren 2017). This non-committal line is also taken here regarding this grouping, and the individual languages will consistently be referred to simply as Indo-Aryan. This includes also those northwestern Indo-Aryan languages spoken in the region that have not normally been regarded as “Dardic” because of their greater similarities with Indo-Aryan varieties in the lowlands. Kashmiri is, with its more than four million speakers (Koul 2003: 897), the region’s largest Indo-Aryan language.¹ Most

¹ Hindko (northern) and Pahari-Pothwari are spoken on the lower slopes of the mountain system and are closely related to, and form a continuum with, some of the large languages spoken by some 100 million people in the lowland south of the Hindu Kush.
of the other Indo-Aryan languages are spoken by relatively small speaker communities, often without official recognition or a standardized written language.

Iranian is the region’s second largest phylogenetic component. The Iranian languages present in the Hindu Kush belong to three separate major groups: Southwest Iranian, represented by various communities that speak a local variety of Dari or Tajik (essentially regional varieties of Farsi); Southeast Iranian, represented by a single minor community, Parachi; East Iranian, represented on the one hand by Pashto, and on the other hand by languages belonging to the Pamir group. These latter consist of a number of mainly small communities with a long-standing presence in the northern-most part of the region (in Afghanistan and Pakistan, as well as in adjacent areas of Tajikistan and China). The regionally most influential of the Iranian languages, and with the highest number of speakers, is (Northern) Pashto, with a geographical spread far beyond this region, in large parts of Afghanistan and Pakistan.

Nuristani (in total only 5–6 languages) is spoken in a system of remote valleys in a rather confined area in northeastern Afghanistan, to some extent spilling over into adjacent Chitral valley in Pakistan. Balti, found in the eastern parts of Gilgit-Baltistan (under Pakistani control), Purik, a closely related language or variety, on the other side of the so-called Line-of-Control, and Ladakhi, spoken in the northern part of the union territory of Ladakh (under Indian control), are all three part of a Northwestern Tibetic grouping, at the geographic fringe of the Sino-Tibetan world. The only two Turkic languages with a relatively stable presence in the northern-most part of the region, Kyrgyz, in the Pamir corridor, and (Southern) Uzbek in Afghan Badakhshan, belong to two separate branches of Turkic. Finally, Burushaski, while having been the subject of numerous attempts at connecting it to various linguistic phyla, is for all practical purposes still best seen as a language isolate. It is spoken by less than 100,000 people in the extreme North of Pakistan, close to the border of China.

Multilingualism is ever-present across the region, but manifests itself locally in different ways. The coexistence of a local, non-dominant language, and at least one dominant language can be observed almost everywhere. Typically, the dominant language, serving as an inter-community communication language, is a national or administrative language, such as Urdu–Hindi (in Pakistan and India), Persian (in Afghanistan), and to an extent English. Those have mainly been acquired through regular classroom instruction. But also, languages with slightly lower status, usually acquired in non-formal settings, such as Pashto in the southwestern parts of the region, or Khowar in Pakistan’s Chitral district, are locally influential and important for cross-community interaction. It is not unusual for an adult man, almost regardless of educational level, whose first language is the language of a small speaker community, to be able to communicate reasonably
well in 3–4 languages, spanning one or more domains. Less seldom, at least until recently, have these more dominant languages succeeded in penetrating the home domain (Liljegren and Akhunzada 2017: 66).

While the region for quite some time has been referred to as particularly interesting in terms of areality and language contact (Emeneau 1965; Masica 1991: 43, 2001: 259; Skalmowski 1985), and a number of features have been suggested as characteristic (Baart 2014; Bashir 1988: 392–420, 1996, 2003: 821–823, 2016; Èdel’man 1980, 1983: 35–59; Fussman 1972: 389–399; Tikkanen 1999, 2008; Toporov 1970), relatively little detailed and systematic areal-linguistic research on comparable data has been carried out so far.

2 Kinship terms and linguistic typology

The study of kinship systems and kinship terminology has a long history, especially in the field of anthropology. These were first studied in a systematic way by Lewis Henry Morgan (1818–1881). Morgan’s resulting distinction between descriptive and classificatory systems was ground breaking, but is in several respects outdated and has been replaced by other typologies (Godelier et al. 1998: 9). His taxonomy of a few basic patterns, subsequently labelled as Hawaiian, Iroquois, etc. (Trautmann 2001: 270–275), was followed by many studies within various subfields of anthropology and linguistics, not least within the American anthropological school (Franz Boas, Alfred Kroeber, etc.). These studies often emphasized the connection between terminological distinctions and family structure, marriage patterns in particular (Nikolayeva 2014: 11). Kroeber (1909: 78–79) identified eight distinguishing features of cross-cultural and cross-linguistic validity (generation, lineal vs collateral, relative age, sex of relative, sex of speaker, sex of connecting relative, blood relatives vs. affinal relative, and condition of life), later evaluated from a cross-linguistic perspective by Greenberg, who concluded that only a few of them were in fact universal (1966: 87).

Most of Morgan’s original assumptions regarding kinship systems and their capability of characterizing entire groups, or indeed families of mankind (Trautmann 2008: 6–7), have been abandoned, and during the last century, kinship terminology has repeatedly been deemed unsuitable as an anthropological study object (Trautmann 2000: 563). Yet, the structural build-up of the kinship domain in various language communities, and how that is reflected in language-specific terminology, has continued to be of interest, and from the 1980s and onward the field has experienced a renaissance of sorts (Godelier et al. 1998) and an alliance with other disciplines is explicitly called for (Trautmann 2001: 268–269), in many ways echoing Lévi-Strauss’ mid-twentieth century conviction that modern
linguistics was bound to play an instrumental role in the study of kinship if any real progress was to be made at all (1963: 31–54). The shortcomings of many earlier anthropological studies were perhaps not so much in their premature mapping of terminologies to the socioeconomics and marriage patterns observed in specific ethnolinguistic groups, as their failure in taking historical evidence, particularly deep historical evidence, into account (Kryukov 1998: 295; Trautmann 2000: 566, 2001: 283–284; Parkin 1990: 70–71).

A few aspects of kinship terminology, and how the kinship domain is structured, are of particular interest from a linguistic – and more specifically from a typological – perspective: First, the durability of such systems is almost unparalleled. Even after radical lexical changes have taken place, or an entire society has gone through major sociological transformations, or long after a language has ceased to be spoken, terminological distinctions remain. That has been observed for instance in regard to the so-called Dravidian terminology in South India, a system whose existence can be traced back, by written records, more than a thousand years (Trautmann 2000: 564). Second, it is a semantic domain that easily lends itself to the study of large-scale evolutionary trends and the development of lexical differentiation (Allen 1989: 184), fully comparable to the cross-linguistic investigation of the (considerably less complex) domain of colour, as pioneered by Berlin and Kay (1969). For kinship, as well as for colour, it has been possible to postulate an overall direction of change (Kryukov 1998: 313; Parkin 1990: 72). Third, the way in which the kinship domain is carved up reflects, in spite of its overall complexity, a limited number of organizing principles, and can (as noted already by Morgan and his contemporaries) be classified as essentially belonging to one of a handful of “systems”. Those systems are in their turn strongly areal in their geographical distribution in much the same way as particular word order typologies are dominant in certain world regions and virtually unattested in others (Trautmann 2001: 282). While “Dravidian” is typical of South India and as such also occurs in non-Dravidian languages in the region (Trautmann 2000: 564–565), it is practically unknown, as a kinship system, in languages of Europe or the Middle East (Trautmann 2001: 282).

Turning to what is known about kinship terminologies in the world region where the Hindu Kush is situated, the Indian subcontinent is relatively well-represented in ethnographic and anthropological literature. Roughly speaking, the subcontinent hosts three different kinship systems, each associated with one of the major language families found here: Dravidian, Indo-Aryan and Munda (Trautmann 2000: 564–566). The much-studied, and already mentioned, Dravidian system is most clearly represented in the languages of South India and Sri Lanka, such as Tamil and some other, mainly Dravidian, languages. Typically, one’s father’s brother is equated with one’s own father, and one’s mother’s sister with one’s mother. A paternal uncle older than one’s father is a “big father” and a younger one is referred to as “little father”. The maternal aunts are likewise “big”
and “little mothers”. Consequently, the children of those fathers and mothers are one’s siblings and thus unmarrigeable. This is often referred to as same-sex sibling merger. On the other hand, cross-cousins, i.e. the children of one’s father’s sister and one’s mother’s brother, are as non-siblings marriageable (Trautmann 2001: 270). The Indo-Aryan (or more correctly North Indian) system, as represented in for example Hindi–Urdu but found in a large number of languages in the populous regions of northern India, Pakistan and Bangladesh, clearly discriminates between parents and their siblings, and it lacks any contrast of cross and parallel cousins. It is primarily structured by the opposition of wife-givers and wife-takers (Trautmann 2000: 565). It is essentially a Muslim culture of kinship, modelled on the Arabian kinship system “as revised and disseminated by Islam” (Trautmann 1981: 109). While the Dravidian and the North Indian systems and terminologies account for the large majority of the population of the subcontinent, the system labelled Munda provides a crucial piece in the puzzle, as a configuration associated with the third major linguistic phylum indigenous to the subcontinent, the Austroasiatic phylum. Like Dravidian, this system is of a so-called symmetric prescriptive type in that it differentiates between cross and parallel kin, and it reflects repeating marriage alliances between groups. Contrary to Dravidian, cross cousin marriages are not allowed (Parkin 1985; Trautmann 2000: 566). Communities characterized by the North Indian kinship system are the only ones of those three that are geographically adjacent to the Hindu Kush. Also highly relevant as far as the Hindu Kush is concerned are those studies that deal with kinship in Himalayan groups that speak Tibetan, or Tibetan-related, languages. Here we have little evidence in the form of written sources and much less of comparative work to lean on, but research points in the direction of an ancestor language of a symmetric prescriptive kind, not unlike what we find in Munda (Allen 1976).

While kinship and social organization has been the object of some geographically more confined studies, e.g. Parkes’ work on the Kalasha (1983, 2003) and Jones’ on Nuristan (1974), there is very little in terms of cross-linguistic research on kin terminology per se with special reference to the mountainous Hindu Kush region, situated as it is at the very outskirts of the subcontinent. Parkin (1987) discusses kin classification in Burushaski and makes comparisons with a few neighbouring languages and comes to some highly relevant conclusions that we will have reason to come back to in the present study, but the truly areal perspective is limited in his work due to lack of comparable cross-linguistic data. A few suggestions are also made regarding kinship terms and Hindu Kush areality in Liljegren (2017: 146–148) and Koptjevskaja-Tamm and Liljegren (2017: 216–217).

Partly building on Kroeber’s feature approach, but reflecting more recent ethnographic work, Nikolayeva (2014: 69) provides a useful framework for studying kinship terms from a modern typological perspective, as one of many
potential semantic domains. She does so, both in terms of morphological structure and semantic structure. The present study primarily makes use of her semantic classification by means of five cross-linguistically relevant features: kinship line, direction of kinship, age, generation, and sex.

3 Sample and data collection

With the aim of obtaining a tight sample, first hand data was elicited from as many of the varieties spoken in the region as was practically possible. This was carried out in a series of multilanguage collaborative workshops held in the period 2015–2018 in Islamabad, Kabul, Fayzabad and Srinagar, supplemented with a few individual sessions in Islamabad, Gilgit and Kargil. The resulting sample, including 59 varieties, is shown in Table 1.

1–3 speakers per variety were invited as native speaker consultants. They were presented with a questionnaire including 95 kin relations represented by descriptive terms in English (e.g. FATHER’S BROTHER’S WIFE) and with direct translational equivalents in one of the regional lingua franca, Dari, Pashto, or Urdu. After having had the task explained to them, they were asked to fill in the questionnaire for their own language, only giving reference terms (not address terms) actually used in the local community, by applying a written representation of their own choice (which in most cases was a modified form of an Arabic-based orthography that they were most familiar with). When they were ready for it, and after a general check for accuracy, the list was voice recorded. All the lists were subsequently transcribed into IPA, based on the recordings aided by the consultants’ written representations. In citing individual examples, the resulting broad phonetic transcription has been used, even for languages for which a phonemic representation has been used in previous publications. The aim with this is to facilitate maximum comparability across the varieties. In some cases, I have chosen to disregard individual pieces of data due to what appear to be obvious mistakes, misinterpretations or translation effects.

4 Kin classification and kin terminology in the languages of the Hindu Kush

Following Nikolayeva (2014: 69–115, 171), five main features by which kinship meaning is lexically expressed cross-linguistically were considered: kinship line, direction of kinship, age, sex and generation. All varieties included in the present
data set were analysed and classified for the presence and prevalence of those features. Kin terms are non-typical lexical items in that they are relational by definition, meaning that an individual is identified and referred to only in relation to another individual. Therefore, in order to avoid ambiguity, I will henceforth use the terms referent and anchor, as defined by Dahl and Koptjevskaja-Tamm (2001:

2 This Nuristani language is also known as Kalasha, a name preferred by many of its speakers, but in order to avoid unnecessary confusion with the Indo-Aryan language with the same name, I have kept the designation Waigali for the present purpose. In the most recent version of the language catalogue Glottolog (Hammarström et al. 2021), these two languages are referred to as Nuristani Kalasha and Chitral Kalasha, respectively.

<table>
<thead>
<tr>
<th>Language Varieties</th>
<th>Language Varieties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burushaski (2)</td>
<td>Burushaski, Hunza [bsk (h)] Burushaski, Nagar [bsk (n)]</td>
</tr>
<tr>
<td>Indo-Aryan (33)</td>
<td>Bateri [btv] Kalkoti [xka] Pashai, Amla [psi (am)]</td>
</tr>
<tr>
<td></td>
<td>Broskat [bkk] Kashmiri, Ind [kas (i)] Pashai, Aret [zee (at)]</td>
</tr>
<tr>
<td></td>
<td>Damel [dml] Kashmiri, Pak [kas (p)] Pashai, Chalas [zee (ch)]</td>
</tr>
<tr>
<td></td>
<td>Gwarbati, Afg [gwt (a)] Khowar [khw] Pashai, Korangal [zee (kg)]</td>
</tr>
<tr>
<td></td>
<td>Gwarbati, Pak [gwt (p)] Kohistani Shina [plk] Pashai, Sanjan [gth (sn)]</td>
</tr>
<tr>
<td></td>
<td>Gawri [gwc] Kundal Shahi [shd] Pashai, Shemal [zee (sh)]</td>
</tr>
<tr>
<td></td>
<td>Gojri, Afg [gju (a)] Pothwari [pfr] Sari [sdf]</td>
</tr>
<tr>
<td></td>
<td>Gojri, Pak [gju (p)] Palula [phl] Shina, Gilgit [slc (p)]</td>
</tr>
<tr>
<td></td>
<td>Hindko, north [hno] Pashai, Alasai [psa (ai)] Shina, Gurez [slc (i)]</td>
</tr>
<tr>
<td></td>
<td>Indus Kohistani [mvy] Pashai, Aliningar [psa (ar)] Torwali [trw]</td>
</tr>
<tr>
<td></td>
<td>Kalasha [kls] Pashai, Alishang [gth (ag)] Ushojo [us]</td>
</tr>
<tr>
<td>Iranian (13)</td>
<td>Dari, Darwoz [prs (d)] Pashto, north, Ind [pbus (i)] Wakhi, Afg [wbl (a)]</td>
</tr>
<tr>
<td></td>
<td>Ishkashimi [isk] Pashto, north, Pak [pbus (p)] Wakhi, Pak [wbl (p)]</td>
</tr>
<tr>
<td></td>
<td>Munji [mnj] Rushani [sgh (r)] Yidgha [ydg]</td>
</tr>
<tr>
<td></td>
<td>Parachi [prc] Sanglechi [sgy]</td>
</tr>
<tr>
<td></td>
<td>Pashto, north, Afg [pbus (a)] Shughni [sgh (a)]</td>
</tr>
<tr>
<td>Nuristani (6)</td>
<td>Ashkun [ask] Kataviri, east [bshe (e)] Prasun [prn]</td>
</tr>
<tr>
<td></td>
<td>Kamviri [xvi] Kataviri, west [bshe (w)] Waigali2 [wbk]</td>
</tr>
<tr>
<td>Sino-Tibetan (3)</td>
<td>Balti [bft] Ladakhi [lb] Purik [prx]</td>
</tr>
<tr>
<td>Turkic (2)</td>
<td>Kyrgyz [kir] Uzbek, south [uzs]</td>
</tr>
</tbody>
</table>
201–202). If I say *My sister just called*, I identify a particular individual by relating her to myself. In this case, my sister is the referent and I am the anchor, i.e. the individual from whose perspective the referent is a sister.

### 4.1 Kinship line

A basic distinction in kinship is that between lineal and collateral relatives. Lineal relatives are either the anchor’s direct ancestors (parents, grandparents, etc.), or the anchor’s descendants (children, grandchildren, etc.). Collateral relatives are all other kin, e.g. cousins and parents’ siblings. In the present classification (using the values fully present, partially present and absent), the first ascending and the first descending generations were considered when determining the presence of a contrast between lineal and collateral.

Lexical differentiation between direct and collateral lines is found in the systems of all languages in this sample. In as many as 49 of them, this feature is expressed fully, which means that the anchor’s parents are distinguished from their siblings just as the anchor’s own children are distinguished from the anchor’s siblings’ children: Sanjan Pashai (Indo-Aryan) *baabaa* ‘father’ – *mambuu* ‘father’s brother’; *puuter* ‘son’ – *bəawtuur* ‘brother’s son’. Partial presence, found in the remaining 10 languages, is for instance manifested when kinship line is considered in the first ascending generation but is neutralized in the first descending generation: Kalasha (Indo-Aryan) *aaja* ‘mother’ – *naana* ‘mother’s sister’; but *təh* ‘daughter; brother’s daughter’. Partial presence of kinship line characterizes five of the six Nuristani languages in the sample, four of the Indo-Aryan and Sino-Tibetan Purik.

<table>
<thead>
<tr>
<th>Feature value</th>
<th>Number of varieties with the displayed value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line absent</td>
<td>0</td>
</tr>
<tr>
<td>Line partially present</td>
<td>10</td>
</tr>
<tr>
<td>Line fully present</td>
<td>49</td>
</tr>
</tbody>
</table>

### 4.2 Direction of kinship

Direction specifies the nature of kinship by one of the intermediate relatives, for instance to what extent there is a lexical differentiation between e.g. grandchildren
based on whether the referents are the anchor’s son’s children or the anchor’s
daughter’s children. Terms of four generations (first and second ascending, and
first and second descending) were considered in classifying the sample languages:
parents’ siblings (whether mother’s or father’s sibling), grandparents (whether
mother’s or father’s parent), nephews/nieces (whether a sister’s or a brother’s
child), and grandchildren (whether a son’s or a daughter’s child). Each language
was classified in regard to this feature as: fully present, partially present or absent.

The feature of direction is partially or fully present in a large majority of the
sample languages. Only in two of them, Yidgha (Iranian) and Pakistani Wakhi
(Iranian), is it altogether absent. To be classified as a language where this feature is
fully present, relatives need to be distinguished in all four generations: Hindko
(Indo-Aryan) daadaa ‘father’s father’ – nāāqāā ‘mother’s father’, tcaatcaaa ‘father’s
brother’ – maamaa ‘mother’s brother’, patrija ‘brother’s son’ – paandzaa ‘sister’s
son’, poontaa ‘son’s son’ – dheetraa ‘daughter’s son’. Nine of the languages where
this feature is fully present are Indo-Aryan and primarily found in the southeastern
part of the region. The only non-Indo-Aryan languages with this feature fully
present are Kyrgyz (Turkic) and Eastern Kataviri (Nuristani). If such differentiation
is missing in 1–3 of these four generations, the feature is considered as partially
present: Dameli (Indo-Aryan) baloojii ‘father’s mother; mother’s mother’, but
p’aaapi ‘father’s sister’ – nan ‘mother’s sister’, nawāāsi ‘son’s daughter; daughter’s
daughter’, but braasāā zu ‘brother’s daughter’ - īspaśi ‘sister’s daughter’. Partial
presence characterizes 46 of the languages. In Yidgha (Iranian), this feature is
altogether absent: maamo ‘father’s mother; mother’s mother’, kookoo ‘father’s
sister; mother’s sister’, nuweso ‘brother’s daughter; sister’s daughter; son’s
daughter; daughter’s daughter’.

<table>
<thead>
<tr>
<th>Feature value</th>
<th>Number of varieties with the displayed value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direction absent</td>
<td>2</td>
</tr>
<tr>
<td>Direction partially present</td>
<td>46</td>
</tr>
<tr>
<td>Direction fully present</td>
<td>11</td>
</tr>
</tbody>
</table>

4.3 Age

The kinship feature age is expressed when two distinctly different terms are used
for differentiating relatives of a single generation by relative age. In the present
classification (using the values fully present, partially present and absent), only
the first ascending and the anchor’s own generation were considered.
The feature of age is displayed in a more diverse way in the region. In 23 of the varieties it is absent, while partially present in 22, and fully present in another 14. In order for this feature to be counted as fully present, a language must distinguish lexically (not merely by adjectival modification of a generic term) between older and younger siblings (in relation to the anchor) as well as between older and younger siblings of the anchor’s parents (although not necessarily between all of them). The latter is the case in Palula (Indo-Aryan): bheeq ‘sister’ – keeki ‘older sister’ – kuutci ‘younger sister’, and meeci ‘mother’s sister’ – geedjeei ‘mother’s older sister’ – lhookjeei ‘mother’s younger sister’. This is also the case in another eight of the Indo-Aryan languages, all spoken in the central parts of the region. It also characterizes Darwazi (Iranian), Kyzgyz (Turkic) and all three of the sample’s Sino-Tibetan languages. If only present in one of the two generations, it has been considered as partially expressed. That is the situation in all of the Nuristani sample languages, 11 of the Indo-Aryan languages, the two Burushaski varieties and in three of the Iranian languages.

<table>
<thead>
<tr>
<th>Feature value</th>
<th>Number of varieties with the displayed value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age absent</td>
<td>23</td>
</tr>
<tr>
<td>Age partially present</td>
<td>22</td>
</tr>
<tr>
<td>Age fully present</td>
<td>14</td>
</tr>
</tbody>
</table>

4.4 Sex

The kinship feature sex is expressed when relatives of one particular generation are distinguished by sex, i.e. female versus male kin. Terms of five generations (first and second ascending, the anchor’s own generation, and first and second descending) were considered in classifying the sample languages: parents’ parents, parents, siblings, children and grandchildren. Each language was classified in regard to this feature as: fully present, partially present or absent.

The feature of sex is present (either fully or partially) in the kinship systems of all of the sample varieties. It is fully present in 45, and partially present in another 14. When the feature distinguishes between male and female kin in the second ascending generation (Prasun (Nuristani) waaw ‘father’s father; mother’s father’ – qiet ‘father’s mother; mother’s mother’), the first ascending generation (Ashkun (Nuristani) dej ‘father’ – araj ‘mother’), in the anchor’s generation (Afghan Pashto wroor ‘brother’ – choor ‘sister’), in the first descending generation (Afghan Uzbek (Turkic) woqel ‘son’ – qiz ‘daughter’) as well as in the second descending generation (Shughni (Iranian) neboos ‘son’s son; daughter’s son’ – nebees ‘son’s
daughter; daughter’s daughter’), it is considered fully expressed. Sex may, however, be neutralized in one of the generations; in this region, full neutralization is limited to the second descending generation: Hunza Burushaski amis ‘son’s son; daughter’s son; son’s daughter; daughter’s daughter’. Apart from in the two Burushaski varieties, partial sex differentiation is found in six of the Iranian languages, two Nuristani languages, and in four of the Indo-Aryan languages.

<table>
<thead>
<tr>
<th>Feature value</th>
<th>Number of varieties with the displayed value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex absent</td>
<td>0</td>
</tr>
<tr>
<td>Sex partially present</td>
<td>14</td>
</tr>
<tr>
<td>Sex fully present</td>
<td>45</td>
</tr>
</tbody>
</table>

An additional differentiation related to sex has been suggested, namely distinctions based solely on the anchor’s sex (=ego’s sex in Nikolayeva 2014), i.e. even when the sex of the referent remains the same. In the present data set, the possible evidence is limited to a lexical differentiation between ‘a female anchor’s brother’s wife’ and ‘a male anchor’s brother’s wife’ on the one hand, and ‘a male anchor’s sister’s husband and ‘a female anchor’s sister’s husband’ on the other hand. Such differentiation is present in 23 of the varieties, and absent (or inconclusive) in 36. Examples: Torwali (Indo-Aryan) dzamaaš ‘sister’s husband (male anchor)’ – saran ‘sister’s husband (female anchor)’. Dameli (Indo-Aryan) braazej ‘brother’s wife (male anchor)’ – zaamili ‘brother’s wife (female anchor)’. However, these marginal occurrences in the Hindu Kush data are in no way comparable with the expression of such differentiation in Nikolayeva’s sample, in which it is applied in relation to close or relatively close consanguineal kin (2014: 92).

<table>
<thead>
<tr>
<th>Feature value</th>
<th>Number of varieties with the displayed value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anchor’s sex absent</td>
<td>36</td>
</tr>
<tr>
<td>Anchor’s sex present</td>
<td>23</td>
</tr>
</tbody>
</table>

4.5 Generation

The feature generation is considered expressed when relatives of different generations are lexically distinguished, whereas non-distinctiveness is described as generational merging. For this feature, two separate realizations of differentiation/merging were considered: consanguineal and affinal. For generational
differentiation in the consanguineal realm, the analysis was limited to second ascending versus second ascending generation of the anchor, and first ascending versus first descending generation lineal kin. For the generational differentiation in the affinal realm, the analysis took three generations of marital relations into account: first ascending (anchor’s spouse’s parents), zero (anchor’s spouse, spouse’s siblings, siblings’ spouses), and first descending (anchor’s children’s spouses).

The feature of generational differentiation is expressed consanguineally in all of the languages in the present sample. This is taken to mean that no merging takes place between any terms referring to the anchor’s grandchildren and any terms referring to the anchor’s grandparents, and that no merging takes place between any terms referring to the anchor’s children and the anchor’s parents. Only in two of the languages, Purik and Balti, both Sino-Tibetan, some merging takes place between terms referring to in-laws of different generations (parents of the anchor’s spouse, spouse or siblings of a spouse and an anchor’s children’s spouses): Purik ama ‘mother; wife’s mother; husband’s mother’ – numu ‘wife’s sister’, but nammaa ‘wife; son’s wife’.

<table>
<thead>
<tr>
<th>Feature value</th>
<th>Number of varieties with the displayed value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generation merging (affinal and consanguineal)</td>
<td>0</td>
</tr>
<tr>
<td>Generation merging partially present (affinal)</td>
<td>2</td>
</tr>
<tr>
<td>No generation merging</td>
<td>57</td>
</tr>
</tbody>
</table>

There are other, more peripheral, generational neutralizations occurring, such as between the anchor’s grandchildren and the anchor’s siblings’ children (Yidgha (Iranian) numis ‘son’s son; daughter’s son; brother’s son; sister’s son). Those have not been considered diagnostic for the presence or absence of the feature generation per se, but will be of importance for further studies.

4.6 Lexification patterns

A few lexification patterns, partly relating to polysemy, have been observed that are of particular interest from an areal perspective. These have to do with (a) terms referring to both of the anchor’s parents, (b) terms referring to the anchor’s grandparents, (c) terms referring to the anchor’s grandchildren, and (d) terms referring to the anchor’s parents’ siblings. For a complete list of those kinship terms in the sample languages, see supplementary material.
Terms referring inclusively to both parents in the data are of two kinds: either a special lexical term is used that bears no clear relation to the kin terms for father or mother, or, more commonly, a co-compound is used in which the individual terms for father and for mother can be recognized. Kinship terms of the first type are only found in 10 of the varieties under study (Brokskat (Indo-Aryan) \(pʰæmæ\), ‘parents’ – boo ‘father’ – aaj ‘mother’; Darwazi (Iranian) \(χuunawaada\) ‘parents’ – baaba ‘father’ – uuma ‘mother’), while as many as 48 of the 59 sample varieties give evidence of co-compounds. For one of the varieties, data for this particular item is missing. Of the 48 co-compounds, 34 follow the order M–F: Kamviri (Nuristani) nuu-toot; Khowar (Indo-Aryan) nan-tat; Gurezi Shina (Indo-Aryan) maa-malu; Pakistani Pashto (Iranian) moor-o-plaar, while the other 14 follow the order F–M: Wakhi (Iranian) tat-nan; Afghan Uzbek (Turkic) aata-aanæ; Nagar Burushaski aua-mama; Balti (Sino-Tibetan) ata-aŋo. Those two patterns have a relatively clear geographical distribution (Figure 1), with most of the F–M compounds in the northeast and the M–F compounds in the southwest. In a general way, this appears to be in line with the macroareal tendencies reported by Wälchli (2005: 216–218): co-compounds for ‘parents’ occur with high frequency in the eastern parts of

![Figure 1: “Parents”-terms in the Hindu Kush.](image-url)
continental Eurasia and in areas considered residual zones. Although the geographical distribution of the two orders M–F and F–M varies in unpredictable ways, the order F–M is typical of the inner parts of the continent while ‘parent’ compounds with the order M–F is dominant in many parts of South and Southeast Asia.

There are two types of grandparent terms: (a) those that make a two-way differentiation based on the sex of the referent (Bateri (Indo-Aryan) doo ‘grandfather’ – daa ‘grandmother’; Munji (Iranian) paap ‘grandfather’ – maama ‘grandmother’), and (b) those that make a (lexical) four-way differentiation based on the sex of the referent as well as the sex of the anchor’s parent (Kalkoti (Indo-Aryan) daad ‘father’s father’ – deed ‘father’s mother’ – maam ‘mother’s father’ – meem mother’s mother). Only 11 of the varieties make a clear four-way differentiation, all of them Indo-Aryan and spoken in the south (Figure 2).

Terms referring to grandchildren display a greater degree of differentiation, but taking the presence of two features into account we can identify three main categories: (a) those that make no differentiation at all (only one term meaning grandchild, Parachi (Iranian) nawaa; Pakistani Wakhi (Iranian) napis), (b) those
that make a differentiation based solely on the sex of the referent (Sawi (Indo-Aryan) nawaasu ‘grandson’ – nawaasi ‘granddaughter’; Ashkun (Nuristani) nuwu ‘grandson’ – nuut ‘granddaughter’), (c) those that make a differentiation that take both the sex of the referent and the sex of the anchor’s child (who is the parent of the referent) into account. The latter does not necessarily mean a four-way differentiation (such as in Pakistani Kashmiri (Indo-Aryan) pootor ‘son’s son’ – pootir ‘son’s daughter’ – djuutor ‘daughter’s son’ – djuutir ‘daughter’s daughter’). Instead, there are various ways in which the two features are combined (for example in Sanglechi navas ‘son’s son; son’s daughter’ – wədər ‘son’s daughter; daughter’s daughter’ – wədaszaat ‘daughter’s son’). Only 10 of the varieties do not make any differentiation at all, while 47 take the sex of the referent into account, and out of those, 15 also distinguish a grandchild of an anchor’s son and a grandchild of an anchor’s daughter. The geographical distribution is similar (Figure 3), although not identical, to the one found with grandparent terms, with the most differentiated terms concentrated in the south. In this case, however, it also includes a few non-Indo-Aryan varieties: Kyrgyz (Turkic) and the Iranian languages Afghani Wakhi, Sanglechi, and Darwazi.

**Figure 3:** Grandchildren terms in the Hindu Kush.
Terms referring to the anchor’s parents’ siblings are of three main types in this region: First, we have a system in which there is full lexical differentiation when referring to the anchor’s father, the anchor’s father’s brother and the anchor’s mother’s brother on the one hand, and full differentiation between the anchor’s mother, the anchor’s mother’s sister and the anchor’s father’s sister on the other. This means that there are six separate terms for these referents, here exemplified by Indo-Aryan Ushojo (Figure 4). This type of terminology is also typical of the northern parts of the subcontinent as a whole. In the Hindu Kush, it is primarily evidenced in the (mainly Indo-Aryan) languages in the South, bordering on the lowland, and, again, in the (mainly Iranian) languages of the far North, while it is largely missing in the more central parts of the region.

![Figure 4: Parents and their siblings: full differentiation (Ushojo, Indo-Aryan).](image)

Second, there is a system in which the anchor’s father’s brother can be referred to with the same term as the anchor’s mother’s brother, and/or the anchor’s mother’s sister is referred to with the same term as the anchor’s father’s sister. I will henceforth refer to this as an “uncles-and-aunts” terminology, here exemplified by Iranian Wakhi as spoken in Pakistan (Figure 5).

![Figure 5: Parents and their siblings: “uncles-and-aunts” terminology (Pakistani Wakhi, Iranian).](image)

Third, there is a system in which the anchor’s father’s brother can be referred to with the same term as the anchor’s own father (usually with the addition of a
modifying ‘big’ or ‘little’), and/or the anchor’s mother’s sisters are also mothers, “big” or “little” depending on their relative age. This I will refer to henceforth as a “mothers-and-fathers” terminology, here exemplified by Nuristani Waigali (Figure 6). Indeed, this mothers-and-fathers type is found in as many as 20 of the varieties, tightly followed by the full-differentiation type, in 19 of the varieties, and finally, 16 of the varieties are of the uncles-and-aunts type.

However, there are a number of hybrid systems and variations within these main types. Therefore, it makes more sense to display the presence/absence of a “mothers-and-fathers” terminology (Figure 7) separate from the presence/absence of an “uncles-and-aunts” terminology (Figure 8) rather than treating individual varieties as belonging exclusively to one type. In e.g. Palula (Indo-Aryan), the terminology is strikingly similar, even form-wise, to the full differentiation displayed in Figure 4 (pʰee, piti, baabu, jeei, meeci, mamaa), but within the community, the usual way of referring to the anchor’s paternal uncles is as gaaɗbaabu (=BIG FATHER) and hookbaabu (=LITTLE FATHER), respectively, and geedjeei and hookjeei for the anchor’s maternal aunts, just like in the terminology of Figure 6. In some of the varieties, a “fathers” terminology is used for paternal uncles, but there is no corresponding “mothers” terminology applied to maternal aunts. In Ladakhi (Sino-Tibetan), abac’senmo (cf. aba ‘father’) is used for the anchor’s father’s older – or perhaps oldest – brother, but for a younger paternal uncle, an altogether different lexeme, agu, is applied, as is the case with the other siblings of the anchor’s parents: ane ‘father’s sister’; azəŋ ‘mother’s brother’; matson ‘mother’s sister’ (cf. ama ‘mother’).

In Nuristan and adjacent areas in the borderland between Afghanistan and Pakistan, some varieties display what seems like a hybrid of the systems represented in Figures 5 and 6. In e.g. Kalasha (Indo-Aryan), only paternal uncles are referred to as gara daada (=BIG FATHER) and tɕʰuutje daada (=LITTLE FATHER), whereas
maternal and paternal aunts are equally referred to as *naana*, contrasting with *aaja* ‘mother’.

An additional polysemy pattern (in addition to its general “fathers-and-mother” terminology), only evidenced in the two Burushaski varieties in this sample, is a term common to an anchor’s father’s sister and an anchor’s mother’s brother, *nana*, i.e. the two opposite-sex siblings that are not considered “fathers” or “mothers” (Figure 9).

In spite of various alternations and hybrids, as discussed above, the three main types with reference to parents and parents’ siblings have obvious sub-areal distributions. A full (six-way) differentiation is mainly a feature of the southern and the northern fringes of the region. The fathers-and-mothers type appears in a number of languages in a central belt, all the way from Nuristan in the west, across the highlands of the region, to Baltistan and Ladakh in the east (see Figure 7). The uncles-and-aunts type overlaps to some extent with the fathers-and-mothers type in the western-most part of the region, i.e. in the highlands north of Kabul valley, but is further to the east confined to the Wakhan corridor and some areas immediately adjacent to it (see Figure 8), and clearly distinct from the “mothers-and-fathers” belt to its south.

**Figure 7:** Distribution of father/mother-polysemy in the Hindu Kush.

"Mothers" and/or "fathers"

- **Absent**
- **F=FB or M=MZ**
4.7 Overall kin typology

In the analysis of her global sample, Nikolayeva (2014: 96) arrives at a classification into eight “language types” (I–VIII) based on the presence (whether fully or partially) and particular combination of the above-treated basic kinship features.
(see Table 2). The feature anchor’s sex is treated somewhat differently in that all the languages in which this feature is present, regardless of the presence or absence of any of the other features, constitute a type of their own. Some of the types (I, II, IV and possibly VIII) are only represented by a single or a few, often relatively small, languages. The features line, generation and sex are present in the overwhelming majority of the languages in the global sample. Among the more widespread types, III is well-represented in Europe, and by some of the world’s major languages; V is represented by major languages of South and West Asia, such as Hindi and Arabic; VI is represented by Japanese and some languages of continental Asia; and VII by a number of languages in Central and East Asia.

Table 2: Language types based on the presence/absence of differential semantic features in kinship terms, following Nikolayeva (2014: 96), and corresponding types evidenced in the Hindu Kush sample.

<table>
<thead>
<tr>
<th></th>
<th>Line</th>
<th>Generation</th>
<th>Age</th>
<th>Dir</th>
<th>Sex</th>
<th>Anchor’s Sex</th>
<th>Example language</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>Tagalog</td>
</tr>
<tr>
<td>II</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Rapanui</td>
</tr>
<tr>
<td>III</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>English</td>
</tr>
<tr>
<td>IV</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>Maori</td>
</tr>
<tr>
<td>V (HK Type 2)</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>Hindi</td>
</tr>
<tr>
<td>VI (HK Type 3)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>Japanese</td>
</tr>
<tr>
<td>VII (HK Type 1)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>Chinese</td>
</tr>
<tr>
<td>VIII</td>
<td>+/-</td>
<td>+/-</td>
<td>+/-</td>
<td>+/-</td>
<td>+/-</td>
<td>+</td>
<td>Alyawarra</td>
</tr>
</tbody>
</table>

Applying this classification on the Hindu Kush sample, three of those types occur: First, those that display all five differential features (line, direction, age, sex and generation). This is Nikolayeva’s VII; I will refer to that as type 1 in the Hindu Kush context. Type 1 is represented by 32 of the varieties, i.e. a majority of the data set. Second are those that display the four differential features of line, direction, sex and generation. This is V in Nikolayeva’s classification; I will refer to that as type 2. Type 2 is represented by 23 of the Hindu Kush varieties. Third, those that display the four differential features of line, age, sex and generation. This corresponds to Nikolayeva’s VI; I will refer to this as type 3. The latter is only represented by two varieties. For another two of the varieties, we lack conclusive data to classify them into one of those types. As can be seen in Figure 10, type 1 is found in the central-eastern part of the Hindu Kush, whereas type 2 is most common in the western part of the region.
Apart from those features, a more general feature of combining or differentiating blood kinship and kinship by marriage is also discussed by Nikolayeva. About half of the Hindu Kush varieties express combinations in one way or another. Ishkashimi (Iranian) ama ‘father’s sister; father’s brother’s wife’; Ladakhi (Sino-Tibetan) kasa ‘older brother; husband’s brother; wife’s brother; sister’s husband; daughter’s husband’s father’.

In an attempt at feature aggregation, taking both the five general features and the specific lexification patterns discussed above into account, 23 binary features were set up (see supplementary material for details) and a NeighborNet analysis conducted. The result can be seen in Figure 11.

While this visual representation is indeed open to a variety of interpretations, we can identify at least two tight clusters, one (at the right-hand side of the figure)
including the varieties from Dameli (Indo-Aryan) to Hunza Burushaski, and another (at the left-hand side of the figure) including the varieties from Rushani (Iranian) to Yidgha (Iranian). The rest of the languages can be considered non-clustered (although some of them cluster at a lower level, usually those that are already geographically adjacent as well as genealogically closely related). Taking those two clusters, hereafter named cluster 1 and cluster 2, and displaying their respective geographical distribution (Figure 12), an interesting pattern appears that suggests significant areal diffusion in two discrete yet geographically (mostly) contiguous zones. One major zone stretches through almost the entire region, forming a central belt from east to west. This zone is phylogenetically diverse, including Indo-Aryan, Nuristani, Sino-Tibetan and Burushaski alike. Another one is situated north of the former, essentially comprising the Wakhan corridor, including fewer languages (exclusively Iranian and Indo-Aryan), but it may also be linked to the Pamir area of Tajikistan, considering the inclusion of Rushani (Iranian) in this cluster, although we lack comparable data to say for certain whether that is the case.
5 Other linguistic features and their geographical distribution

There are good reasons to consider the feature clustering described in the previous section as yet another piece of evidence for at least one diffusion zone – although a continuously disintegrating one – extending through this mountainous region, setting it off from the Indo-Gangetic plain in the South, the Central Asian steppes in the North as well as the large Persian-dominated areas to the West, and possibly linking it in a significant way with the rest of the Himalayan region to the East. While most purely morphological or syntactic properties appear to be of less interest in this regard, a few structural (mainly semantically grounded) features, indeed point in the same direction as the distribution of kinship terminologies.

The dominant numeral base in the Hindu Kush is vigesimal, i.e. counting in 20s. This has been pointed out in the past (Bashir 2003: 823; Èdel’man 1983: 50–51; Koptjevskaja-Tamm and Liljegren 2017: 220–221; Tikkanen 1988: 309), and is also confirmed in the data discussed here, in which 38 out of the 59 varieties include (although always in combination with a decimal component) a vigesimal base.
According to this system, the numeral ‘36’ is structured as e.g. $20 + 10 + 6$ (Hunza Burushaski altʰer-turna-micindo). The geographical distribution is one of a continuous central belt, stretching through the region from East to West (to a large extent overlapping with the presence of a “fathers-and-mothers” kin terminology or an “uncles-and-aunts” terminology), sandwiched between consistent decimal systems in the northern-most languages and in a southern belt bordering the lowlands. In a northern subarea of the vigesimal region, partly coinciding with the extension of the “uncles-and-aunts” terminology, numeral composition consistently follows the linear order base $+ n$ (e.g. $20 + 10 + 6$, as in Kalasha (Indo-Aryan) bići-ze-daz-e-ʂowa; and Yidgha (Iranian) ysto-les-uṣo), while in a southern cluster it is consistently $n + \text{base}$ (e.g. $6 + 10 + 20$, as in Ushojo (Indo-Aryan) şi-e-ga-bi), with a hybrid structure, $n + 10/20 + n$, in an area between the two (e.g. $20 + 6 + 10$, as in Alishang Pashai (Indo-Aryan) wast-u-tʂ-ui).

Other shared lexical patterns or what seems like a strikingly similar way of organizing individual semantic domains have been identified: fine-grained and lexically distinct calendrical expressions; multi-dimensional and multi-degree deictic contrasts cross-cutting a number of part-of-speech categories; and the prevalence of co-lexicalized intensifiers (Koptjevskaja-Tamm and Liljegren 2017: 220–223). All of these need a more systematic study based on a larger data sample.

Structural differentiations related to animacy (contrasting animates and inanimates and/or humans and non-humans) is another property shared by many of the languages in the Hindu Kush region (Bashir 2003: 823; Èdel’mann 1983: 46), but expressed most clearly along a SW–NE axis (Liljegren 2019: 310–319). In several varieties of Pashai (Indo-Aryan) – in the present data sample evidenced in Alingar, Amla, Chalas, Aret, Korangal – and in Dameli (Indo-Aryan) it co-occurs with sex-based gender differentiation, mainly with copulas as targets. In Kalasha (Indo-Aryan) and Khowar (Indo-Aryan) an animate-inanimate distinction is reflected both in lexical choice and with copulas and auxiliaries as targets; here, however, inherited sex-based gender is altogether lost (Bashir 2016: 267). In Burushaski, an intricate four-gender system is partly based on animacy, partly on sex (Berger 1998: 33–38).

The presence in the region’s languages of retroflex affricates (and/or fricatives) mostly along with a set of retroflex plosives (the latter a much more widespread South Asian feature) has been discussed extensively (Hock 2015; Tikkanen 1988, 1999, 2008; Toporov 1970). Regardless of its origin (whether substratal or as a parallel development or shared retention), it is a feature of well over half of my 59 varieties, represented by Indo-Aryan, Nuristani, Iranian, Burushaski and (marginally, at least in the Hindu Kush region itself) Sino-Tibetan, occurring in a central belt throughout the Hindu Kush. Related to that, and also dependent on it, is the presence of a tripartite dental—retroflex—palatal (or more correctly apico-dental vs. apico-alveolar vs. lamino-alveolar; Strand 2001: 255) affricate/fricative
contrast, in the Indo-Aryan languages reflecting an inherited (but possibly areally reinforced) contrast, lost in Indo-Aryan elsewhere (Bashir 2003: 822).

Finally, another set of properties, although not reflected in the majority of Hindu Kush languages, cluster in a significant way in the western-central part of the region, coinciding with the habitat of those Nuristani and Indo-Aryan communities that until very recently adhered to a pre-Islamic religious world view and cultural expressions – and indeed still do in the case of the non-Muslim Kalasha (Indo-Aryan) – which with slight local variations once may have dominated and characterized the entire mountain region. Among these, the presence of retroflex/rhotized vowels (alternatively approximants, depending on analytical preferences), pronominal kinship suffixes, bisyndetic contrast marking and a geomorphic system of spatial reference deserve to be mentioned (Di Carlo 2011; Heegård and Mørch 2004; Liljegren and Svärd 2017; Heegård and Liljegren 2018). In the case of retroflex vowels, it is tempting to draw parallels with the articulatorily similar Burushaski phoneme that Berger (1998: 22) describes as “ein stimmhafter retroflexer Sibilant mit gleichzeitiger palatal-dorsaler Engebildung” [a voiced retroflex sibilant with concomitant palatal-dorsal closure]. Also, the extension of the geomorphic coding found in the languages of Nuristan and lower Chitral may continue on northeast-ward and thus include Burushaski (Berger 1998: 92–94; Heegård and Liljegren 2018: 154).

5.1 Kinship terms, language contact and shared cultural values

The geographical distribution of certain kinship features/types in the Hindu Kush, as observed in Section 4, and its striking overlap with other areal-linguistic findings in the region, as noted in Section 5, begs for an explanation. In the following we will take a brief look at reasonable clues as far as cultural traits – particularly in marriage patterns – and intercommunity relations are concerned.

Today, most cultural influences flow from the dominant political centres in the South, on respective sides of the international borders, often transmitted through languages of wider scope, such as Urdu and Dari. However, there is little evidence of any large-scale interaction going on between the lowland cultures of South Asia and the high-altitude areas of the Hindu Kush until relatively recently. Instead, the ties between the various local communities remained stronger, probably for centuries, than any of those that arose from time to time between individual communities and parts of lowland society (Cacopardo and Cacopardo 2001: 26, 31). The non-Muslim section of the small Indo-Aryan Kalasha community is today the remnant carrier, virtually in the entire Hindu Kush region, of practices and a
worldview explicitly connected with a pre-Muslim belief system (or rather a spectrum of related religious practices) that once characterized this region, or at least substantial parts of it. Cacopardo and Cacopardo (2001: 249) summarize this former shared worldview as “an articulate and coherent system of meanings, associated with a particular form of human adaptation to the ecosystem”. This in its turn provided “the bases for an ethics, a legal order, a political structure, an architecture of economic relations, a system of technological and ritual prescriptions”. From a historical perspective, Jettmar (1975: 464, 1982: 256) and Parkes (1987: 649–652) depict a culture centred around goat-herding, with a fundamental opposition between pure and impure, a dualism further reflected in the opposition, or complementarity, between a male sphere and female sphere. The former was associated with the higher regions, inhabited by the fairies and gods, transhumance and livestock, and the latter with the lower regions, i.e., the river beds and the village areas, human contamination (particularly as it relates to childbirth and menstruation in e.g. Kalasha practices) and agriculture. Along with that there was also a strict division of labour (Cacopardo and Cacopardo 2001: 69–70, 249–250; Edelberg and Jones 1979: 50). Other important, and probably ancient, ingredients were hunting traditions, witchcraft and so-called milk adoption (Jettmar 1982: 256; Parkes 2003; Tucci 1977: 69). The last-mentioned term refers to the practice of local rulers to give their children to be fostered by a family in the area in order to form strategic (and intimate) allegiances.³ Politically, there was a pronounced difference between the flat “brothers model” of power sharing and reciprocity governing relations and economy within the local communities, as opposed to the “father model” characterizing the urbanocentric or “civilized” world surrounding (and gradually encroaching on) the area, the latter ruled by hierarchical structures, military force and centralized powers (Cacopardo and Cacopardo 2001: 250). Having said that, it would be totally incorrect to assume that this region, in spite of its relative inaccessibility and self-reliability, was isolated from large-scale developments. For instance, the massive, and often rapid, spread of Islam in large parts of West and South Asia affected the Hindu Kush region, too, and that already at an early date, although often in an uneven way. While

³ That this type of foster-relation was nurtured well into the mid-twentieth century is evidenced by a personal experience from 1998. I was at the time in the process of establishing contact with the Palula community in the Chitral valley, and was taken along with my family by Dr. Bakhtiar-ud-Din, a member of the ul-Mulk family (the ruling dynasty of Chitral until the abolition of the princely state in 1969), to visit Biori, a remote Palula village. During the visit I was pleasantly surprised to be welcomed into a house, and to be seated with both male and female members of the household, something that would otherwise rarely happen in the traditional and strictly (outside the family) gender-segregated Sunni contexts of the Hindu Kush. Our guide, the prince, explained that “these people are my milk brothers and milk sisters”.

Badakhshan (in today’s northeastern Afghanistan) fell under Arab control as early as in the seventh and eighth centuries AD, most of the area to its immediate South (i.e. Nuristan and adjacent areas of Pakistan) was still unconverted as late as two centuries ago (Cacopardo and Cacopardo 2001: 32–39).

This very gradual – and still incomplete – process of Islamization is reflected in intra-regional differences in marriage patterns and local alliance building. Somewhat simplified, there is on the one hand a deep-rooted, largely pre-Islamic, preference for clan or lineage exogamy and small-scale inter-community interaction, evidenced in various communities throughout the Hindu Kush region (Cacopardo and Cacopardo 2001: 42). On the other hand, there is a competing tendency in many of the region’s communities to favour, at least rhetorically, parallel cousin marriages (Parkin 1998: 253), often reinforced by a common Muslim culture with its ultimate origin in the Arabian kinship system (Trautmann 1981: 109). The “original” exogamic pattern is primarily observed among the non-Muslim Kalasha, but it persists, or at least persisted, also among some of their more recently converted neighbours (Cacopardo and Cacopardo 2001: 107). Barth (1956: 83), for instance, observed a “dislike for marriages with close relatives of any kind” in Chitral, and several researchers have reported specific rules against marrying someone related within a set number of generations in pre-Muslim Nuristan (Cacopardo and Cacopardo 2001: 196; Robertson 1974: 86). In some of the communities that currently identify themselves strongly with Islam, e.g. the Dameli, Palula and Shina Kohistani communities, a gradual but significant increase in the number of endogamic – at least valley-endogamic – marriages, can be detected, in contrast to the relatively large circles of marriage ties of previous generations (Cacopardo and Cacopardo 2001: 152–153; Zarin and Schmidt 1984: 51). The competing “Islamic” pattern is particularly strong, and has probably been so for a long time, among the Pashtuns (i.e. speakers of Pashto). In many of the Pashtun-dominated societies, an emphasis on patrilineal parallel cousin marriage co-occurs with the formation of alliances between groups of patrilineal cousins against more distant relatives (Lindholm 1996: 30). This is also a group that has expanded in the Hindu Kush region in the last few centuries, often at the expense of many smaller, primarily Indo-Aryan, communities (Strand 2001: 254). The same pattern, by which parallel cousin marriages are encouraged if not necessarily prescribed, is therefore prevalent also in many of the non-Pashtun groups residing at the southern fringes of the Hindu Kush that are heavily influenced by Pashtun culture, e.g. in Kohistan (Knudsen 2009: 32, 97), but also in other, Muslim-majority, groups of the larger region, e.g. among the Wakhi (Shahrani 2002: 83).

Turning to possible connections between particular marriage patterns and the kinship terminology applied by individual language communities, it is essential to avoid the pitfall (referred to in Section 2) of trying to match synchronic (or near-
synchronic) cultural observations with semantic patterns without realizing that
the semantic organization may reflect temporally very distant social realities.
Parkin (1987, 1998), who is well aware of this issue, makes some crucial obser-
vations regarding Burushaski kinship terminology. He concludes (1987: 165) that
the prescriptive terminology (i.e. the lexi-
fi
cation pattern that I referred to as “fa-
thers-and-mothers” terminology in Section 4) must have been the ancestral ter-
nomology of Burushaski, not altogether different from the Dravidian ones of South
India, while the neighbouring variety of Indo-Aryan Shina, in its essential parts
adhering to the same terminological differentiations, moved in the direction of
Burushaski, not the other way around (Parkin 1987: 167). The latter is evidenced by
the fact that Indo-Aryan in general prefers a system of full differentiation (i.e. a
non-prescriptive system), like in Hindi–Urdu and many of the major languages of
Northern India and lowland Pakistan. What Parkin was not aware of at the time
(1987: 166), was that the Shina pattern is not an isolated phenomenon. Instead, as
we saw above (Figure 6), it is a pattern reflected one way or another in as many as
20 of the 59 sample varieties (representing four of the six genera in the Hindu
Kush), all occurring in a central west-to-east-stretching belt across the entire region.
While Burushaski cannot possibly be postulated as the (only) source of this
lexification in all of those languages, some of them spoken very far away from
today’s Burushaski-speaking areas, the general scenario painted by Parkin (1990: 74), is still valid: The population speaking a non-Indo-European language (in this
case Burushaski) shifted to an Indo-European language (in this case Shina) but
retained the pattern of the original language. This is in line with observations
elsewhere, in which the prescriptive pattern of a substratal language is retained
when shifting to a non-prescriptive language, such as with Dravidian populations
shifting to Indo-Aryan in Southern India (Allen 1989: 183). The only other neigh-
bouring languages for which Parkin had comparable data were Indo-Aryan
Khowar and Iranian Wakhi, and he correctly concludes that these two languages
do not display the same lexification pattern (Parkin 1987: 167). Khowar, along with
some of its Iranian neighbours, displays a diametrically different pattern (with its
“uncles-and-aunts” terminology), which is neither typically North Indian nor
Dravidian, and again a very atypical one for an Indo-European language. However,
the same type of historical language-shift scenario as with Burushaski and Shina
can be applied, although the identity of the substrate in this case remains un-
known. That Khowar displays such a non-Indo-Aryan pattern makes perfect sense
if we consider Morgenstierne’s (1932: 51) hypothesis that the common ancestor of
Khowar and Kalasha represents the earliest northward migration of Indo-Aryans
into this region. For a prolonged period of time it must have been a relatively minor
component in an area where non-Indo-Aryan (most likely now entirely lost) lan-
guages dominated (Parpola 2002: 92–94; Tikkanen 1988), at the time isolated from
the rest of Indo-Aryan. It is, however, reasonable to assume that groups of speakers of some of those other languages shifted to a predecessor of Khowar-Kalasha once it became a more influential element in its new environment.

It is therefore hardly surprising that we find what appear to be the oldest accessible layers of areality in the inner, and more remote and high-altitude, parts of the Hindu Kush. While most of the descriptions of marriage patterns referred to above are tied to the western part of the region, where Islam still is a relative newcomer and where dramatic cultural change came much later, we have reason to believe that the entire “crescent” (corresponding to the two clusters in Figure 12) once was a contiguous diffusion area of sorts, virtually forming a string of multiple small-scale cross-community links, reflected in a cluster of linguistic micro-areas (Tikkanen 2008: 258). The area stretching from Nuristan in the Southwest, across the Burushaski area in the North, to the Tibetan-dominated areas in the Southeast would perhaps not primarily have been defined in terms of cultural and linguistic homogeneity or shared features across its entire distribution. Instead it might have been a significant refuge zone, with several local centres of shared retention, innovation and diffusion, and probably with a high degree of phylogenetic diversity prior to Indo-Iranian penetration (Tikkanen 2008: 258), much in line with Nichols’ (2003: 306) characterization of the Caucasus. These highland regions of Central Asia could very well have comprised several languages or language families as distinctly different from one another as Burushaski is today vis-à-vis the other surrounding languages of the Hindu Kush (Tikkanen 1988: 321). Some of those are likely to be extinct, without leaving any traces save for sporadic substratal effects, such as the ones we noted for Khowar. Other languages co-existing or spoken adjacent to them would definitely have been Sino-Tibetan, making up a component of the Karakorum and the Hindu Kush just as ancient as Burushaski (Tikkanen 1999: 146). Although not a phylum normally associated with this part of the subcontinent, a previous presence of Austroasiatic in Himachal Pradesh has been postulated to explain some lexical and structural properties of Indo-Aryan as well as Sino-Tibetan languages spoken in that area, i.e. not too distant from the Hindu Kush (Tikkanen 1988: 319). Although it is unlikely that Dravidian languages have ever been spoken as far north as in the Hindu Kush, archaeological evidence points to cultural contacts between the pre-Aryans of Swat (in the Hindu Kush) and the presumably Dravidian-speaking Indus civilization. Similarly, there is archaeological evidence connecting the (pre-Aryan) Neolithic population of Kashmir (perhaps responsible for some peculiar, and clearly non-Indo-Aryan substratal effects on Kashmiri) and Swat with the East Tibetan Plateau (Tikkanen 2008: 258; Xu 1991: 22).
6 Conclusions

In the present areal-typological investigation of kinship terms in the Hindu Kush region, I conclude that the geographical distribution of a few differentiating properties of language-specific kinship terminologies in most respects are in line with the distribution of a number of other, previously identified, mainly lexico-semantic, properties. Significant geographical clustering along a belt in the region’s inner, and more remote, parts, vis-à-vis properties largely associated with the surrounding spread zones, is interpreted as evidence of a remnant ancient contact zone.

Particular attention was paid to a few notable polysemy patterns, especially the terms used for parents and their siblings. Two significant subareas were identified, a smaller northern “non-differentiating” belt, collapsing the terms for one’s paternal and maternal uncles and aunts, and a larger central “selectively differentiating” belt, in which paternal uncles are “big” or “little” fathers, and maternal aunts are “big” or “little” mothers. Those patterns are taken as evidence of more than one substratum underlying many of the Indo-Iranian languages. In the latter case, Burushaski or a Burushaski-related language certainly plays a significant role, but also other phylogenetic components (without any remaining presence) are likely to have shaped the current distribution of kinship terminologies and types found in the region.

Acknowledgements: A special thanks to all the speakers of various Hindu Kush languages who contributed valuable first-hand data, and to the Forum for Language Initiatives (Islamabad), Samar (Kabul and Fayzabad), and the Department of Linguistics, University of Kashmir (Srinagar) for excellent collaboration in facilitating workshops and other elicitation sessions. I would also like to thank Noa Lange, Stockholm, for assistance in processing and annotating the audio recordings. I am also thankful to John Peterson, Kiel University, for insightful suggestions and help in processing feature aggregation and the visualization of structural distances. Thanks also to the editors and two anonymous reviewers for many helpful comments and corrections.

Research funding: This work is part of the project Language contact and relatedness in the Hindukush region, supported by the Swedish Research Council (421-2014-631).

References


**Supplementary Material:** The online version of this article offers supplementary material (https://doi.org/10.1515/lingty-2021-2080).