LINGUISTIC WATERSHEDS: A MODEL FOR UNDERSTANDING VARIATION AMONG THE TIBETIC LANGUAGES

Brad Chamberlain
Payap University
<brad_chamberlain@sil.org>

Abstract
This study applies the observation of alignment between geographical watersheds and linguistic groupings to the Tibetan Plateau and the Himalayas. Tournadre (2014) estimates 220 Tibetic language varieties in 25 major groupings, sharing a common linguistic ancestry. Typological groupings can be readily identified through mapping human settlements to watersheds. For areas that have yet to be researched, consistent hypotheses for typological groupings can be arrived at. Next to explaining anomalous data within a particular area or how certain linguistic features spread, a watershed-based map identifies possible linguistic areas to be researched. The concept is applied in detail to the watersheds and languages of Bhutan and then expanded out to the broader Tibetan region.

Keywords: Bodish, watersheds, variation
ISO 639-3 codes: adx, bod, bro, cgk, dka, dzl, dzo, goe, jul, kgy, khg, kjz, kkf, lep, lhm, lhp, lkhd, luk, neh, npb, ole, scp, sgt, tgf, tsj, twm, xkh, xkz

1 Introduction
The people of the Tibetan Plateau and surrounding mountains speak numerous related language varieties. Tournadre (2014:13) estimates 220 Tibetic language varieties in eight major groupings. These languages share a common linguistic ancestry, while the peoples’ histories, cultures and religious institutions are all interconnected. As a result, there are many related but mutually unintelligible language varieties. In the Tibetan region, typological study and language mapping have proven difficult, not only due to the linguistic complexities, but also due to the logistics of a harsh climate, extreme geography, and political sensitivities. The contours of this rugged landscape affect the contours of the languages. High mountain ridges and raging rivers form barriers for human migration and thereby mark language boundaries. These same river valleys are also the highways through which the languages have spread.

This paper demonstrates that watersheds correlate closely with known linguistic groupings through the region. While it is true that people groups have spread across mountain passes and between multiple watersheds, typological groupings can be readily identified through mapping human settlements to watersheds. This is not a new observation. It is clear to anyone working in the region that mountains and rivers form the contours of human populations and their languages. However, with this paper, I am proposing that in the Tibetan region, the influence of the mountains is so great that it can be used predictively to set expectations regarding unresearched language areas. The extreme nature of the landscape sets a predictive framework for looking at the relationship of the languages in the area. Where the linguistic reality falls outside of the predictive patterns laid out through analyzing the watersheds, further explanations are necessary (e.g., recent migrations, etc.), but these are expected to be the minority of situations, not the majority.
The result is a new view of the relations between the language varieties in the region. For areas that have yet to be researched, consistent hypotheses for typological groupings can be arrived at. A watershed-based map of the languages in this region provides a starting point for understanding the relationships between language varieties. The interesting work can then begin: identifying watersheds which have not been researched, explaining how certain linguistic features spread across major watersheds, or explaining anomalous data within a particular watershed. In this paper, the concept is applied in detail to the watersheds and languages of Bhutan, then expanded out to the broader Tibetan region to underline the theory’s relevance. At the broader Tibetan level, predictions are possible regarding language variation. However, the next necessary step is to corroborate these predictions with typological linguistic data. This paper sets up a hypothesis for what will be found through that typological study, but such a study is outside the scope of this current paper.

It is important to note that at this point, I am presenting a hypothesis and research method. Follow up work, as detailed in section 6, is essential to corroborate and to take advantage of these ideas. My hope is that this paper will shed light on our collective understanding of the relationships among Tibetic languages.

2 Key issues affecting the Tibetic languages and research of those languages

This paper focuses on Tibetic languages, a cover term for the Tibetan Central Bodish language varieties as well as some of the surrounding related language varieties, such as the East Bodish languages of central Bhutan, and Tshanglaish, including Southeast Bhutan, western Arunachal Pradesh. East Bodish languages are clearly the most closely related to Central Bodish of these languages, while the Tshanglaish languages are linguistically less similar, but have a very strong overlay of vocabulary due centuries of a shared Tibetan Buddhist identity, as well as other linguistic exposure.

The same geographic and political features which lead to the language variation and sociolinguistic issues in the region are also part of why there are still significant numbers of unresearched language varieties in the area. In this section, we will look at some of the key factors which are at play in the Tibetic linguistic region.

2.1 Geography

Tibetic languages are spoken in settlements from as low as 290 meters (i.e., Phuntsholing, Bhutan) above sea level, up to 4870 meters (i.e., Wenquan, China). The area has high mountains and deep valleys. Villages less than a kilometer apart on opposite sides of a river may be more than a days’ journey away from one another. Similarly, villages on opposite sides of mountain ridges may be near each other as the crow flies, but many days from one another as people walk. This geographic variety leads to isolation and variation among the language groups in the area. It also makes access to many of the language areas difficult.

2.2 Politics

The Tibetic languages are indigenous to five nations: China (where they are found in five provinces: Qinghai, Tibetan Autonomous Region, Sichuan, Gansu and Yunnan), India (where they are found in five states: Jammu & Kashmir, Himachal Pradesh, Uttarnachal, Sikkim, and Arunachal Pradesh), Pakistan (in the Baltistan region), Nepal (most of the northern third of the country) and Bhutan (throughout). Many of these language communities live in or adjacent to disputed territories between these countries. As a result, the already heightened political sensitivities of this meeting point of world cultures are even more pronounced. These political restrictions not only impact the isolation of the language groups, but also keep many groups outside the limits where direct linguistic or anthropological research is possible.
2.3 Layers of linguistic interrelations
The Tibetic languages have close linguistic affiliation. The Central Bodish languages share common ancestry from a form of Tibetan very similar to the Classical Tibetan preserved in written Buddhist scriptures of the region. The East Bodish languages of Bhutan are related to the others as well, but are more divergent, likely splitting from Central Tibetan at a point prior to the writing of Classical Tibetan. Tshanglaish languages are more distantly related to the others.

In addition to shared linguistic origins, there have been centuries of intermixing due to migrations back and forth between regions and due to a common religious language and religious infrastructure, including a common honorifics system across the region. Both linguistic and cultural mixture has occurred over the centuries. The result is a tangle of interrelated languages in which two varieties a thousand kilometers apart may have features in common which neither share with neighbors. This intermixing is especially poignant in the Central Bodish languages. Tournadre (2014:13) states:

“This incredible diversity, which is largely due to the extension of the Tibetic linguistic area and to the geographic isolation of many localities, has been underestimated until quite recently. As mentioned earlier, some of these languages are tonal, while others are nontonal. Although they do share a common basic vocabulary and grammar, they may largely differ in some aspects of phonology, morphology, syntax and lexicon. In general, the languages listed above do not allow mutual intelligibility. However some of them do allow a certain degree of intelligibility.”

Language variation exists at many levels, but seems particularly notable in lexicon, phonology, tone systems, numbering systems, morphology, and in verb phrase phenomena (particularly issues such as evidentiality and mirativity). As an example, consider the following plots of lexemes (Figure 3) for the word
‘hair’, which are all clear cognates of the Old Tibetan \textsuperscript{55}skra, and of the local lexemes for ‘village’, where there is more variety, although similarities exist across large regions.

\textbf{Figure 2.} Cognate forms of ‘hair’ throughout region, but variation in forms of ‘village’. ‘Village’ has cognate forms across large areas, but also variation at some local levels

2.4 Diglossia
A key issue in understanding the language varieties of this region is the diglossic relationship between the spoken, or profane, languages and the Classical, or holy, language. Speakers of many of the Tibetan varieties view their own language as a corrupted variety of the high language. Most people may not have mastery of the classical language, but they still view it as a pure form of their own language. From this perspective, these are all one language, with different spoken varieties. Yet those spoken varieties contain significant variation and many are unintelligible to one another. As a whole, these languages can be seen as a complex system of interrelated language varieties.

On the Tibetan Plateau, the diglossic link is strong, and mutually unintelligible spoken forms are commonly viewed as varieties of a singular language. Thus, on the plateau, there may appear to be significant under-differentiation between language varieties, as compared with the non-plateau Tibetic areas. Off the plateau, both to the east in China (through the mountains of the ethnic corridor, where Khams Tibetan mingles with the Qiangic languages) and to the south in the Himalayan language areas, diglossia tends to be weaker at varying degrees. In these areas, the people are likely to see themselves as ethnically distinct from “Tibetans.” Nationalistic movements in many areas push this even further. As a result, there may be what appears as over-differentiation of language varieties in these regions, as compared with the language variation on the plateau.

The result is that language varieties on the plateau which are considered part of the same language may have more internal variation than there is between southern varieties considered as separate languages. These conflicting ideals of what constitute a language, combined with the unity these languages have through their diglossic relationship to Classical Tibetan, are also complicating factors in accurate language mapping and classification in this region.

2.5 Current enumeration of language varieties
It is not in the scope of this current paper to present a typological study of Tibetan languages. Rather, it is to lay out a paradigm for accurate and efficient study of these languages. To demonstrate the effectiveness of watershed analysis, section 4 of this paper will demonstrate typological and linguistic data within Bhutan, as a subset of the Tibetic language area. For the broader region, I defer to Tournadre (2014), who has the most complete published categorization of the Tibetic languages to date.

Tournadre (Ibid:16) breaks the Tibetic languages into eight groupings, not including the East Bodish or Tshanglaish languages. In the article, he lays out some basic typological patterns and demonstrates how some of the key languages distinguish from one another. These linguistic patterns group geographically. “The classification proposed here is essentially based on a genetic approach, but it also includes geographical
parameters, migration and language contact factors. The languages... can be grouped together at a higher level into 8 major sections. Each section constitutes a geolinguistic continuum. These are: North-Western section (NW); Western section (W); Central section (C); South-Western section (SW); Southern section (S); South-Eastern section (SE); Eastern section (E); North-Eastern section (NE).” These groupings are shown in Figure 4. Within those eight groups, he cites 50 main “languages” and over 200 varieties, roughly shown in Figure 5 (Ibid: 13).

**Figure 3. Tournadre’s eight major groupings**

![Map showing eight major groupings](image1)

Tournadre looks at the geographic proximity in this analysis. Languages are grouped based on linguistic similarity, but also on geographic proximity. Section 5 of this paper demonstrates that watersheds are the geographical feature which account for the groups of related Tibetic language varieties. The watersheds, in fact, provide both a more detailed picture of which language varieties are likely to group together, but also demonstrate the mechanism for why the languages have developed in this way.

**Figure 4. Tournadre’s 50 main languages. Note that there are many more spoken Tibetic language varieties not plotted on this map**

![Map showing 50 main languages](image2)

3 Watersheds as a part of a biolinguistic system

3.1 Watersheds

A **Watershed** is the area that drains to and is drained by a river and its tributaries. Watersheds often encompass several cultural and national boundaries. What happens in one country’s part of the watershed will impact water quality, quantity, and the people who depend on it in the countries downstream.
In the Tibetan region, the high mountain ridges form the boundaries between numerous watersheds, river systems which flow from the high altitudes down to the lowlands. The Tibetic region is the source of many of the world’s largest river systems, including the Yellow river, the Yangtze, the Mekong, the Salween, the Tshangpo, the Indus, the Sutlej, the Ganges, and the Brahmaputra. These rivers are the direct water sources for over 200 million people and indirectly support more than a billion people. These watersheds are the byways through which people have moved about for centuries. In these high mountain regions, the ridges between watersheds are difficult to cross at best and impossible to cross in many cases. Human migration occurs regularly along the watersheds such as occasional movements over accessible passes and as isolated movements over more difficult mountain passes.

The barriers between these watersheds are surmountable in some cases on the plateau and in the very distant lowlands, far outside of the Tibetic language regions and in domains where other language groups and families are dominant. Most of the Tibetic languages exist in areas where only occasional mountain passes, accessible during short intervals each year, allow transit between watersheds. As a result, populations remain, for the most part, tied to particular watersheds. Because of this, it is possible to map populations within watersheds as a predictive tool in order to form starting hypotheses regarding languages spoken in each watershed within the Tibetic region.

That watersheds form the main byways of human migration is nothing new (cf., Davison 2006 and Coblin 2002). However, the implications of this fact seem to be both significant and underutilized in Tibetic language research.

### 3.2 Migration mechanism: Cost-distance model

Cost-distance is a concept related to businesses in which the shipping of goods is necessary. Rather than focusing on the physical distance between a source and destination, cost-distance tools “determine the shortest weighted distance (or accumulated travel cost) from each cell to the nearest source location. These tools apply distance in cost units, not in geographic units.” (ArcGIS Resources 1999).

This same principle applies to human migration. A person is more likely to move along a pathway which expends less energy and brings with it less risk. As a result, human interaction along a watershed across hundreds of miles may be more likely than human interaction between populations physically near to each other, but on opposite sides of a major physical barrier. This basic calculation is at the heart of migration patterns of people and is particularly marked in the Tibetan region’s extreme geography.

### 3.3 Speciation and linguistics

Mechanisms for linguistic variation have analogies found in biological speciation. Looking at linguistic groupings through watersheds lends itself to an analogy with biological speciation along geographic lines. A basic understanding of these biological processes is useful in understanding the linguistic processes which lead to language variation in this environment. Klinkenborg (2012) states:

> “Human languages evolve far more quickly than the single species — *Homo sapiens* — that speaks them. And languages speciate for some of the same reasons that organisms do — topographic separation, for instance. A good example is New Guinea, which is as rich in linguistic species (972 endemic languages) as it is in biological species. Topographically speaking, New Guinea is famously difficult, a torturous landscape that isolates humans as well as other species.”

Speciation is the process where a new species develops as a result of geographic, physiological, anatomical, or behavioral factors that prevent previously interbreeding populations from breeding with each other. In the case of languages, the analogue would be the process where a new language variety develops as a result of geographic factors that prevent interaction between communities whose languages, over time, develop into mutually unintelligible languages.

There are three types of biological speciation which are relevant to the discussion: peripatric speciation, parapatric speciation, and hybridization.
PERIPATRIC speciation occurs when small populations of organisms become isolated in a new environment. Speciation occurs in the newly isolated, smaller population. The FOUNDER EFFECT causes rapid speciation due to lack of contact with the source population. Linguistically, as members of a language community move into a new watershed (perhaps by migration over a mountain pass during summer months), they then settle into a new watershed. They are now cut off for the most part from the source population. Language variation occurs in isolation from the source language population and is likely to undergo rapid change. Migration between large-scale watersheds which do not have a confluence in the Tibetic region will likely lead to the peripatric analog in linguistic change.

PARAPATRIC speciation occurs when a small population breaks off from a main population into a new habitat, but there is no physical separation between the two populations. Rather, it is still probably that there will be contact between the groups on a regular basis, but for the most part they are separated. The language analogy would be language communities which settle into isolated valleys within a watershed, or beyond transit ravines in the watershed, or on opposite sides of a major river in a watershed. In this case, the populations live separately for the most part, and language change occurs. However, the smaller community continues to have some contact with the majority community, so the language varieties continue to develop some aspects in common as well.

A third relevant form of speciation is HYBRIDIZATION. Hybridization occurs when compatible species are in proximity with one another, and occasionally, in biology, reproductive isolation is achieved, leading to separate species. In biology, hybrid zones exist where the ranges of two interbreeding species meet and cross-fertilize. These areas are likely to have a wide variation of species and only occasionally lead to new species. Linguistically, this is analogous to language contact issues. Unlike biology, languages continue to mix, even if they are not genetically related. Hybrid zones would find their analogy at two levels.

1. Areas where related language varieties mix regularly. This could be at a confluence of watershed systems into a common watershed, where the languages of the watersheds are related. In such an area, the language variety in the catchment area will likely share features with more divergent language varieties upstream. The internal variety in that CATCHMENT LANGUAGE would be great as it incorporates many influences from related varieties. As such, though, this variety would also have an impact on converging the languages upstream as well.

2. At a broader level, as languages enter new watersheds (particularly as the rivers move further down from the plateau), they are more likely to have contact with very different languages. Hybridization will occur in such areas as features from the different languages in contact mix with one another.

These three forces, peripatric speciation, parapatric speciation, and hybridization are important biological analogies for the language phenomenon which occur in the watersheds of the Tibetic region. Together, they describe the key mechanisms underlying watersheds as geographic markers of language variation.

3.4 Biological diversity and linguistic diversity
Another area of relevant analogy between the biology of watersheds and the linguistics of watersheds is found in studies which correlate biological and linguistic diversity. Gorenflo (2012) states:

“Results indicate that these regions (of high biological diversity) often contain considerable linguistic diversity, accounting for 70% of all languages on Earth. Moreover, the languages involved are frequently unique (endemic) to particular regions, with many facing extinction. Likely reasons for co-occurrence of linguistic and biological diversity are complex and appear to vary among localities, although strong geographic concordance between biological and linguistic diversity in many areas argues for some form of functional connection. Languages in high biodiversity regions also often co-occur with one or more specific conservation priorities, here defined as endangered species and protected areas, marking particular localities important for maintaining both forms of diversity.”
The same geographic features which lead to biological variation also lead to significant linguistic variation. Further, with more than a billion people depending on the water from this region, the watersheds are political hotbeds. What happens to the watersheds affects not just the populations within the high mountains, but all of the populations down river. At the same time, the languages in each watershed and at various elevations are interacting with different flora and fauna. This leads to diversity in lifestyle, in housing, and in culture, and to varied vocabulary to accommodate these different realities. This biolinguistic view of the watersheds of the Himalayas should give us cause to work towards preservation and revitalization of both the linguistic and biological systems which inhabit these watersheds.

4 Putting the theory to work: Bhutan as a testing ground

Bhutan is located in the Himalayan mountains, surrounded by India and China. Bhutan’s elevations range from 97 meters (318 feet) above sea level up to 7570 meters (24,836 feet). All of this altitude variation is within only 145 kilometers (90 miles) north to south, and 306 kilometers (190 miles) east to west.

![Figure 5](image.jpg)

*Figure 5. Bhutan in relation to the Tibetic-speaking region, and geographic map demonstrating the extreme geographical variation in this small country*

4.1 Language orientation to Bhutan

According to Van Driem (1991), there are 17 languages indigenous to Bhutan. These languages fall into four main typological groupings, including Central Bodish, Eastern Bodish, Tshanglaish, and a handful of small older languages, indigenous to this area since before the strong Tibetan influence of the past 1000 years. Some of these smaller languages are occasionally grouped with East Bodish. These basic distinctions in Bhutanese languages are well attested. For examples, see van Driem 1994, 2011 and Bradley 1997.

Figure 7 adds color to van Driem’s original 1991 language map. In this map, the Central Bodish languages are in red, the East Bodish in shades of green and blue, and Tshangla is in purple.
The Bhutanese languages provide a useful subset of data to represent issues in the wider Tibetic region. Whereas most of the Tibetic languages are Central Bodish, the Bhutanese languages include also Eastern Bodish and Tshanglaish languages. Thus, the typological distinctions are easy to sort out. Issues of migration and language contact are more readily apparent in Bhutan than in other parts of the region because of the clear typological variation between these three major classifications. Looking at these linguistic groupings in light of Bhutan’s watersheds, a clear correspondence is evident. This look at Bhutan, then, helps to demonstrate what is likely to be found with further study in the broader Tibetic region.

4.2 Mapping watersheds: Detailed mapping.
Bhutan’s watersheds run north to south. For the most, part they originate within the mountains of Bhutan. However, several rivers have their origins on the China side of the border. All of Bhutan’s watersheds flow south into the Brahmaputra River. Bhutan’s intricate system of rivers and streams can be seen in Figure 8.

Note that the yellow colored river systems in the south (east, central, and west) do not converge within Bhutan into distinctive watersheds. Rather they are the up-river systems of watersheds that coalesce in India. These three regions correlate very closely with the Lhotshampha, Nepali speaking areas of Bhutan. Bhutan’s river systems are portrayed as watersheds in Figure 9.
In this, we can see seven major watersheds flowing through Bhutan: the Amo, Wang and Puna in the west, and the Mangdi, Bumthang, Kuru, and Kulong in the east. Note that these later four converge within Bhutan and collectively make up the Manas watershed. Also note that the Amo (west) has origins in the region of China between Bhutan and Sikkim, while both the Kuru and Kulong also have origins on the China side of the border.

4.3 Discussion of watersheds and languages in Bhutan
Not surprisingly, the watershed map and the language map have similar contours. These similarities are causally linked: the watersheds directly lead to the human migration patterns, resulting in similar linguistic contours and watershed contours.

4.3.1 Correlation between watersheds and languages
Figure 10 shows the Linguistic Survey (van Driem 1991) language boundaries overlaid on the major watersheds.

In this view, it is clear that the language map contours are related to the watershed map contours. (Note that there is some roughness to the original linguistic survey map, so the contours don’t match precisely.) Here are some quickly observable patterns:
1. The Nepali language areas (the yellow areas found below the blue lines in south west, south central and south eastern Bhutan) map quite closely to the areas outside the main watersheds.

2. The Dzongkha language (western region, red/purple) is dominant across the region of the three western watersheds.

3. The East Bodish languages (found through central Bhutan, blue/green) all fall within the Manas watershed. There are no East Bodish languages outside of that watershed.

4. The Tshangla language (eastern Bhutan, colored in purple), which is very similar to Kalaktang Monpa on the Arunachal Pradesh side of the map, spreads across watersheds. In this, the Tshangla language is not adhering to the watershed prediction, which suggests that there is more to the story. It raises the question of which watershed Tshangla originated from. Given the break in the continuity of the East Bodish languages along the Kulong Chhu, it seems likely that the Tshangla language has more recently spread into this area, isolating the East Bodish languages. The fact that Tshangla does not fit the expectations shows that there is a story here which needs further research to tell.

4.3.2 Tightening up the language map
Using the theory that the human populations are linked more closely within watersheds than they are across watersheds, it is now possible to tighten up this language map. Figures 11 shows (1) a rough population map showing where human habitations are within Bhutan and (2) a map showing those populations along the river systems. Figure 12 then shows an overlay of Van Driem (1991) language mapping on the population mapping.

**Figure 10. Populated areas of Bhutan, demonstrating the river systems as the key settlement pattern**

![Figure 10](image10)

**Figure 11. Linguistic survey of Bhutan boundaries overlaid on settlement map**

![Figure 11](image11)

By tightening up the map to outline languages based on where human populations live within these watersheds and language areas, we can arrive at a much more detailed map (Figure 13) regarding the languages in Bhutan.
In this map, only the areas within Van Driem’s language boundaries, which have human populations, are highlighted. (There is an exception in that Nepali speaking areas in southern Bhutan were not plotted for this study). In Figure 14, there is a bit more added detail giving a good idea of what language should be found in each location within Bhutan.

4.4 Discussion of new language map and what can now be observed about Bhutan’s languages

Looking at this new language map, some issues start to become clear regarding the languages of Bhutan. It is much easier now to make some hypotheses about issues such as dialect variation, language shift, and language maintenance.

4.4.1 Eastern Bhutan with a focus on East Bodish languages

In eastern Bhutan, there is also a lot to notice in this map, in particular, the East Bodish languages (highlighted in green and blue shades in Figure 14-18).
First, Khengkha is found at the convergence of the four rivers that make up the Manas watershed. However, the other languages each are found primarily contained within individual watersheds. Thus, Mangdikha and Trongsakha are on opposite sides of the Mangde Chhu, Bumthangkha is on the Bumthangkha Chhu, Chalikha and Kurtoekha are on the Kuru Chhu, and Dzalakha and Dakpakha are found alongup different branches of the Kulong Chhu.

*Figure 14. East Bodish languages of Bhutan and the Manas Watershed*

Though there is variation among the languages in the Manas watershed, they are more similar to one another than they are to languages of surrounding watersheds. As a group, they have developed along the lines of peripatric speciation. However, there are further language distinctions within the individual smaller watersheds, so along the Mangdi, Khengkha, Mangdikha, and Trongsakha are all found. At this level, the language differentiation is less and is more analogous to parapatric speciation.

Khengkha, spoken at the convergence of these rivers and up each of the four rivers, would have interaction with the largest number of other East Bodish languages, making a “catchment language,” as mentioned in section 3.3. With this, we should be a significant amount of hybridization. Khengkha interfaces with Mangdikha and Trongsakha on the Mangdi Chhu, influencing those languages and being influenced by them. It does the same with Bumthangkha and Chalikha as well. This dynamic will almost certainly result in Khengkha being a somewhat central form among these languages and allow for it to accommodate more significant internal variation than might be found in the other languages.

Note, however, that Khengkha is not contiguous to the Kurtoekha language or the languages along the Kulong Chhu. Rather, Kurtoekha and Chalikha are separated by the Central Bodish Chochangacakha, or Kurmedkha, language, shown in red in the center of Figure 16.

*Figure 15. Chocangacakha and its position amidst the East Bodish languages of the Kuru river*
This language is a more recent migrant community from the west of Bhutan, which settled into that region. Looking at the current language map, a hypothesis could be generated that the ancestors of the Chochangacakha speakers migrated to this area and spread over time, cutting Chalikha off from what was likely a contiguous language group with Kurtoekha. There is solid academic research to support this (Gyeltshen 2006), but further study of the typological similarities between Chalikha and Kurtoekha would be useful to further support the hypothesis of their shared origins.

Similarly, the Tshangla language appears to have found a foothold in the Kulong and part of the Kuru watersheds, and as its influence has grown, other east Bodish languages have been separated from one another. This separation leads to language change. Moreover, the influence of the new language contact causes language change through hybridization. Chali, for example, is said to have vocabulary from many of the local languages, including Kurtoe, Tshangla, and Dzongkha. This small language area, without enforcement from other East Bodish languages, appears to be undergoing rapid change and is likely also threatened as these two languages continue to grow in size and influence.

It seems likely that there was a time when the Dzalakha, Dakpakha, and Tawang Monpa (see § 4.4.2) languages, while not spoken by contiguous populations, were spoken in a watershed with no other languages between them. This would have kept them intermixing and influencing each other. However, the entry of Tshangla into this part of the Kulong River seems to have cut the direct link between these East Bodish languages. As a result, they are each more likely to diverge from one another. The smaller, cut off language of Dakpa, seems likely to be threatened.

The upper Kuru River crosses the border into Lhozhag County, of Lhoka Prefecture in the Tibetan Autonomous region. Based on the watershed theory, it is likely that the villages along the watershed in this area speak languages very similar to the East Bodish languages of Bhutan. This would be very valuable to corroborate with data.

*Figure 16. Gongduk and Olekha languages (yellow) amidst the East Bodish languages*

At an earlier historical level, we can hypothesize that the East Bodish languages at some point migrated along the Manas watershed in territories where Olekha (one larger and three small yellow language areas in the southwest of Figure 17) and Gongduk (yellow language area in the south east of Figure 17) were spoken (van Driem 1995). As the East Bodish languages were increasingly successful, the older languages were relegated to smaller territories. Their territories are not linked in a watershed to any other languages which would reinforce them linguistically. Furthermore, they are surrounded by the currently successful and (apparently) growing languages of Tshangla and Khengkha. These two languages also are likely to be threatened.
There are three places at which an East Bodish language appears to have spread over a pass and into another watershed. This spread is fairly recent because significant language change has not yet happened. However, it should be expected that these areas will experience fairly rapid shift. These include the following.

1. The lower Bumthangkha region, where the language spreads out of the Bumthang watershed and over into a tributary of the Kuru watershed, bringing the Bumthang language down towards Khengkha. This actually may contradict the expectation of rapid language speciation because the main road now also goes over this same pass, connecting these two Bumthangkha populations in a way that watersheds never did. It is likely that this modern layer will enable ongoing connection between this Kuru watershed Bumthangkha group and the original Bumthang watershed Bumthangkha-speakers.

2. The Mongar Kheng dialect, spoken along the lower part of the Kuru Chhu, is not contiguous with the rest of Kheng, but is broken by the Gongduk language. Gongduk was in the area first. Khengkha, more recently spread across a pass from the lower part of the Bumthang Chhu, into the Kuru watershed. This Mongar Kheng variety is cut off from the rest of Kheng and is the most divergent of the Khengkha varieties. One could imagine that as Mongar Kheng and the lower Bumthangkha language mingle, a unique regional variety might develop over time. It is also likely that Gongduk will be overtaken by Khengkha, at which point the Mongar Kheng population would be linked back with the main Kheng continuum.

3. Dzalakha also has moved from its place at the north of the Kulong Chhu (we speculate that the Kulong Chhu is its traditional home because it affiliates linguistically more closely with Tawang and Dakpa than with Kurtoekha) and into the Khoma Chhu tributary of the Kuru chhu. It should be anticipated that there will be divergence of these into distinct Dzalakha varieties over time, with Khoma taking on Kurtoe and possibly Chocha characteristics, while Dzalakha is increasingly influenced by Tshangla.

Tshangla, while spreading over the Kulong watershed and to both sides into other areas, makes it into the Kuru watershed, but does not cross the river. The Kuru River marks the current western boundary of Tshangla’s spread. From this, there are many research questions to pursue, but just from looking at the map, along with some basic known information about the languages of this area, we can generate a working hypothesis of how things may have developed in this area. Now data can be gathered to corroborate or contradict this.

4.4.2 A note about extending hypotheses based on watershed predictions.
The headwaters of the Kulong Chhu originate in the south of the Tibetan Autonomous Region of China and flow south through a small part Arunachal Pradesh, in India, and then into Bhutan. The Tawang Monpa language is spoken in this section of Arunachal Pradesh, and the Cuona Monpa language is spoken further upstream in China. Both of these languages are part of the Kulong watershed, and thereby part of the Manas watershed.
Without knowing anything about either of these languages, we can postulate that as they are within the Manas watershed, they will be related to the other East Bodish languages found in this watershed. More specifically, they are likely to be more closely related to the languages in the Upper Kulong river, including Dakpakha and Dzalakha. In fact, linguistic data bears this out (van Driem 2007). In this way, the watershed view of Tibetic languages allows us to predict the typological affiliation of language varieties, and to better conduct targeted linguistic research that confirms or denies these hypotheses. If it confirms the hypotheses, this is support for the approach and allows researchers in the area to use these claims. If not, we must further explore the situation behind that language variety and consider why it is anomalous, what its history is, and where it came from.

4.4.3 Western Bhutan with a focus on Central Bodish languages
Below are some hypotheses about the Dzongkha and other Central Bodish languages based on Figure 14 and isolated in Figure 19.

Dzongkha is found along three major watersheds in western Bhutan. Without knowing anything more than what is presented in this map, we can expect a few major variations in Dzongkha along the three major watersheds, so we can hypothesize dialect variation into three main sections: the Amo, the Wang, and the Puna watersheds.

- In the Amo watershed (the westernmost one), there may be related language varieties found upstream even up into the Tibetan region. Within the populated parts of Bhutan, the Amo Dzongkha is somewhat cut off from the rest of Dzongkha by the Lepcha and Lhokpa communities marked in yellow. They are also up against the Nepali language area, the region marked white in the south west). From this lack of
connection with other Dzongkha-speaking areas, and due to lack of watershed-based language reinforcement, we can hypothesize either language change (i.e., hybridization with surrounding languages) or language shift, particularly to the growing Nepali language. It would not be surprising to find that Dzongkha in this area is not stable.

- On the Wang watershed, the central one of the three Dzongkha watersheds, it seems likely that varieties at the southern end are more similar to varieties to the north than they are with varieties in the Puna watershed, the furthest east of the three. It also may be significant that the Wang breaks into three separate river systems towards the head waters, the Haa Chu, Paa Chu, and Thim Chuu. These would be expected to be more similar to one another than they are to varieties in the Puna or Amo watersheds.

- On the Puna watershed, there is a long chain of settled areas north to south, and we would anticipate some linguistic similarity across this area, yet there are some major tributaries, which would likely show variation from one another. Again these should all be more similar to one another than they are to the Dzongkha spoken in the Wang or Amo watersheds.

At the northern end, there is an unpopulated break before the northern Laya variety. Note there are two other isolated Central Bodish varieties on northern tributaries of the Puna, Lingzhi and Lunana, which are discussed separately from this current Dzongkha discussion. Based on looking at the watersheds of the Dzongkha-speaking area, Figure 20 illustrates predictions of Dzongkha dialects.

**Figure 19.** The Dzongkha speaking areas, including river systems and hypothetical dialects based on watersheds

Given the analogies with speciation, we can expect localized variety in individual small watersheds which share features in common with others in their shared watershed. From this, a taxonomy of the hypothetical Dzongkha dialects is shown in Figure 21.
Figure 20. Hypothetical taxonomy of Dzongkha dialects based on watershed analysis

In this taxonomy—based purely on watershed analysis, not on any linguistic data, further than knowing that these areas speak forms of Dzongkha)—there is a hypothetical expectation of three major groupings of Dzongkha varieties, along the three main watersheds: Amo, Wang, and Puna. These are large watersheds, each made up of multiple tributary watersheds. As such, these original three watersheds could be anticipated to subdivide into six sections, each of which is more linguistically related internally than they are with other groups. Finally, these are subdivided into the individual headwater watersheds, making a potential of sixteen hypothetical Dzongkha varieties.

With this analysis, the next step would be to confirm or deny the hypothesis. This could be done by collecting a set of typological data to compare the varieties at each of these levels and see if this hypothesis is confirmed or if it needs modification. There will certainly need to be modification, which will most likely be along the lines of hybridization due to regularly travelled passes, or relative age of the different watershed settlements, and due to contact with other language areas, particularly with Nepali in the south.

After making this hypothetical projection, I learned of the following list of Dzongkha dialects by Karma Tshering (2013). Tshering lists the following Dzongkha varieties in his Kinship and language documentation in Bhutan: Layap, Lunap, Wang, Punap, Dagap, Chukha, Sha, Hap/Parop, Dranjop. Note that the ‘p’ ending on these is the Dzongkha cognate of the Tibetan –pa, denoting ‘people’, while the name of a district or region is given, so ‘Layap’ is ‘people of Laya’, such as Dranjop, which is Denjongke, or Sikkimese, is outside of the scope of this current discussion. In Figure 22, the first map shows Tshering’s dialect areas. The second shows the above hypothetical taxonomy overlayed with Tshering’s groupings.
These two maps are very similar and suggest that the underlying motivation of the groupings as listed by Tshering is, in fact, the watersheds. The map on the right adds potential further nuance to understanding the dialects of the area, by sub-dividing into headwater watersheds. Moving on from Dzongkha, it is worth looking anecdotally at the three northern Puna watershed varieties. I have not yet seen any linguistic data from any of these varieties. However, anecdotally, within Bhutan, the Laya language is seen to be quite similar to Dzongkha, while the other two are more divergent and have been described to me as being more like nomadic languages from Tibet. So Laya may be linguistically part of the Puna watershed, but due to its separation from other populations, language shift has occurred, greater than that within the contiguous Dzongkha varieties. However, Lingzhikh and Lunakha may be more recent settlements by nomadic Tibetans who crossed the high passes in summer and settled into the previously uninhabited but habitable high land to the south.

In fact, there are five such groups within Bhutan, all of which appear to be related to Tibetan nomads (see Figure 23). These groups all have settled in habitable land just below the glacial meadows where they herd their yaks during the warmer months. It can be assumed that these are separate migrations at various times over the high passes from the north. Having found sufficient grazing ground and a habitable area south of the passes, these populations stayed on. These would be populations mostly cut off from their home population so again, rapid linguistic variation is expected. The groups are readily seen all the way across Bhutan, west to east, with Lingzhikh, Lunakha, Lakha, Brokkat, and Brokpa. Note that Chochangacakha is the result of a migration of Dzongkha speakers, so is not part of the same discussion.
Of these, it is uncertain if Lunakha is a dialect of Dzongkha or if it is a settlement from northern Nomadic groups. Based on cost-distance, it could be either as they are far removed from the Dzongkha speaking population, but isolated from the northern groups by a high pass. Based on current anecdotal evidence, I am grouping with the Nomadic groups, rather than Dzongkha, although Tshering has grouped them with Dzongkha. See Figure 24. Though these five languages likely have common origin, they have been isolated for different periods of time and from one another, so they have all developed to distinctive but related languages. These languages make a great example of an anomaly we might find within the watershed view of languages, which highlight that another process is being overlaid onto the watershed process.

**Figure 23. Proposed taxonomy of Bhutan’s Central Bodish languages**

Based on the watershed viewpoint of this language area, this is a good starting point for understanding the relationships of the Central Bodish languages in Bhutan, including the likely dialects within Dzongkha. This gives a framework for collection of detailed data which will either corroborate or correct this.

**4.5 The method for applying watershed projections to research planning**
Considering the above discussion, this is the method I am proposing.

1. Map detailed watersheds
2. Plot existing linguistic understanding to these watersheds
3. Based on principles of cost-distance and speciation, make hypotheses regarding the gaps in the current linguistic understanding
4. Note where existing data does not fit with these expectations and where this is not current information
5. Research to fill in the gaps in the data and to explain the inconsistencies

**5 Applying at a larger scale – The Tibetic Region**
Section 4 demonstrated the value of looking at the languages of Bhutan through the watershed lens. Now I extend the model to look at the watersheds of Tibet. Refer to section 2.5 above for current language maps of Tibet based on Tournadre (2014). Tournadre’s overview is extremely helpful, and it pulls threads together in a good way across the region. However, it shows what is known, but does not give a view of what is not known nor give us a way to predict what will be found in those unresearched areas. There is a general grouping of linguistic and geographical proximity in the way that Tournadre has developed his eight language groupings. However, there is no underlying explanation of why those groupings exist.
This section will demonstrate that underlyingly, the natural groupings of Tibetan languages are directly along the lines of the watersheds of the region. Watersheds are clearly responsible for most language variation in the region. There is internal similarity between languages within a shared watershed, but more variation with those in other watersheds. By applying this further, we can see (1) what areas are currently unresearched and (2) what the languages spoken in those areas are expected to be like. As we saw in the Bhutan data, there is regular migration between watersheds, back and forth, over the years. Moreover, there is contact with other language areas, so anomalies to the watershed theory will arise. Yet, by mapping the major watersheds of the Tibetan region, we can generate a fairly accurate mapping of the major language divisions in the region. Then with targeted linguistic typology research, we can systematically sort out the variation within those watersheds.

5.1 Mapping watersheds
Figure 25 shows the major river systems of the Tibetan region. These river systems flow from the plateau and down to the east, south and west.

This network of rivers supplies water to one-fifth of the world’s population. In total, there are 7 rivers which flow to the ocean from headwaters in Tibet, including the Yellow River, the Yangtze River, the Mekong, the Salween, the Tshangpo (which becomes the Brahmaputra), the Ganges, and the Indus. In addition, there are significant tributaries to many of these rivers, which also flow from Tibet and later join with those 7 major rivers. These include the Sutlej, which joins the Indus, the Karnali, Gandaki, and Koshi rivers of Nepal which all join into the Ganjes; the Teesta river, along with the aforementioned rivers of Bhutan, including the Amo, Wang, Puna, and Manas, which all flow into the Brahmaputra. There are also significant branches of the Yellow, Yangtze, Mekong, and Tshangpo rivers which create linguistic watersheds.

In Figure 26, these watersheds are highlighted, making nine major watersheds in the Himalayan region: the Yellow, the Yangtze, the Mekhong, the Salween, the Tshangpo, the Indus, the Sutlej, those that join the Ganjes, and those which join the Brahmaputra.
Figure 25. Major watersheds of the Tibetic-speaking region

It is striking to see the narrow east west corridor through which three of these rivers flow, each running north to south. This convergence happens at the south eastern corner of the Tibetan Autonomous Region, and the northwest corner of Yunnan. Here the Yangtze, Mekhong, and Salween flow north to south in watersheds which all flow within a space of only 180 Kilometers (110 miles). Passes between the rivers exist but are mostly over 4,000 meters in elevation and difficult to traverse. This area, referred to as the Three Parallel Rivers, is potentially a great trial ground for the watershed theory. Do languages far north within a watershed have more in common with a local variety than do those just 50 kilometers to the side in another watershed?

In the following description of the region, the correlation between the extreme geography, and extreme biodiversity is noted. Similarly, we should anticipate significant linguistic diversity. According to UNESCO 2003:

“The Yangtze, Mekong and Salween which run approximately parallel, north to south, through steep gorges which, in places, are 3,000 m deep and are bordered by glaciated peaks more than 6,000 m high. The property spans a large portion of the Hengduan Mountains, which is the major arc curving into Indochina from the eastern end of the Himalayas. Being located in the convergent regions of the three world’s major biogeographic realms, the property is an epicenter of Chinese biodiversity. It may also harbor the richest biodiversity among the temperate areas of the world.”

Similarly, these watersheds are home to considerable linguistic diversity among the Tibetic languages found in this region. The Three Parallel Rivers region is a transition zone between Central Tibetan, Southern Kham Tibetan, Eastern Kham Tibetan, and Qiangic languages. It is expected that watersheds will be a useful factor for sorting out the linguistic relations of the many Tibetic language varieties spoken here.
5.2 Tibetic languages and the watersheds

The watershed map has similar contours to the language map shown in section 2.5. Figure 27 shows them together.

*Figure 26. Tournadre’s 50 main languages mapped to watersheds*

In this view, it becomes clear that the groupings Tournadre has developed align very nicely with the watersheds of the area. This is not a surprise as he attributes his groupings to linguistic and geographic factors, and watersheds are certainly a geographic feature. However, looking at the major watersheds provides hypothesis regarding unresearched areas. Mapping is clear for the most part.

**Table 5:1. Tournadre’s Tibetic groupings and their Watershed correlates**

<table>
<thead>
<tr>
<th>Tournadre designation</th>
<th>Watershed</th>
</tr>
</thead>
<tbody>
<tr>
<td>North Western</td>
<td>Indus</td>
</tr>
<tr>
<td>Western</td>
<td>Sutlej</td>
</tr>
<tr>
<td>South Western</td>
<td>Ganjes</td>
</tr>
<tr>
<td>Southern</td>
<td>Brahmaputra</td>
</tr>
<tr>
<td>Central</td>
<td>Tshangpo</td>
</tr>
<tr>
<td>North Eastern</td>
<td>Yellow</td>
</tr>
<tr>
<td>Eastern</td>
<td>Yangtze – western tributary</td>
</tr>
<tr>
<td>South Eastern</td>
<td>Three Parallel Rivers – Salween, Mekong, western Yangtze. Although it may be worth sorting out the differences across these watersheds as well, possibly making more groupings than Tournadre has suggested.</td>
</tr>
</tbody>
</table>

Note that Tournadre’s Southeastern grouping includes languages in the Salween, Mekong, and the westernmost tributary of the Yangtze, i.e., the Three Parallel Rivers region. The eastern group mostly correlates to the Yangtze watersheds other than the far western. Finally, off the northern tip of the Yangtze watershed, the northernmost of the “Eastern” varieties, the Jone, is found in the Yellow watershed.

For the most part, however, the watersheds, as expected, encompass the languages within each of Tournadre’s groupings, but because they extend over a broader region, we are able to assume that other populations within each of those watersheds will speak languages more closely related to others in their watershed than with languages in different watersheds. So a population in far western Tibet, south towards the border is more likely to be like the languages of Spiti and Lahul of India, while languages a bit further north are more likely to be similar to the languages of Ladakh and Baltistan.

Regarding this Western Tibet or Ngari region, Zeisler (2012:3) notes, “For the time being, I do not understand the reason, why the Ngari dialects should be distributed over three language groups.”. The three groups referred to are what she classifies as (1) “Western Innovative Tibetan,” including the languages of the
Sutlej and Indus watersheds, west of Mt. Kailash, (2) “Central Tibetan”, found to the east of Kailash, along the Tshangpho watershed, and (3) “Northern Kham Tibetan”, found in the Ngari region, north of Kailash, in the western-most reaches of the Salween watershed. The Salween doesn’t reach this far, but this area does make up the broad headwaters of that watershed. This quandary is readily explained by the very watersheds I have used to describe the locations of these language groupings within Ngari. The reason “Why the Ngari dialects should be distributed over three language groups” is because they are spread across three watersheds. However, watersheds will account for any further differentiation between the Sutlej and Indus watershed languages, as also suggested in Tournadre’s groupings, rather than grouping them all together as Western Innovative Tibetan.

**Figure 27. TAR watersheds flowing southward to Nepal and Bhutan rather than into Tshangpo**

![TAR watersheds](image)

It is expected that languages in the Lhoke region of the Tibetan Autonomous Region will be found to be similar to the East Bodish languages of Bhutan as these languages also fall within the headwaters of the south-flowing Manas watershed, not the Tshangpo.

There are also languages which are typically grouped with the Central Tibetan oral varieties, such as Gyantze, Tingri and Kyirong, which lay on watersheds which travel south into Nepal, rather than north into the Tshangpo. As such, they may be more appropriately linked with Nepal-based languages to their south: Kyirong to Yohlmo, it’s southern watershed neighbor; Tingri to Jirel and Kagate, found south along their watershed; and Gyantze to Lhomi.

### 6 Typological research proposal

Now data needs to be gathered. Typological data gathered and compared along watershed-based lines will be essential to confirm or deny the theories set forth in this paper. At this point, it is theory and needs corroboration from specific linguistic data. According to the ideas set forth in this paper, watersheds appear to be a powerful predictor of language variation in the Tibetic region. Due to the mechanics of Cost-Distance calculations and principles of speciation, there can be some baseline assumptions:

1. Language varieties within a watershed are more similar to one another than they are with languages in other watersheds.
2. Languages are likely to also diverge along different sub-watersheds or tributaries within a single watershed.
3. Languages at the convergence of watersheds are likely to share elements of languages found upstream in the sub-watersheds.
4. Where watersheds converge outside of the Tibetic region, the links of those rivers are irrelevant to the theory.
5. Where language data does not conform to these expectations, there are other significant mechanisms at play—most likely recent migration—which need to be explored.

The question remains how the languages are similar and dissimilar from one another. Typologically, there are some aspects which are in common between all of these Tibetic languages. But there are also variations. Some of the variations are found throughout languages across the whole region. However, it is expected that some variations correlate to major watershed divisions. Finding these typological markers will
enable us to better understand the key similarities between groups of Tibetic languages, which would constitute the basis for a high level typological grouping. Then, when language data is found in a watershed which does not match the local characteristics, it is possible to understand where this variety has migrated from.

Similarly, within the major watersheds, there are multiple further division of watersheds. The Manas watershed languages will all share one level of typological similarity. But the Mangdi, Bumthang, Kuru and Kulong will each have differences from one another. It should be anticipated that the ways in which Mangdi watershed languages differ from Bumthang watershed languages will be dominant in certain typological areas. This then would allow a second level of typological grouping. Then, further, within the Kulong watershed, there are significant tributaries, each of which may also be found to separate from one another in particular ways, forming yet another typological level.

By collecting a standard set of typological data, it should be possible to define what separates the language varieties in one watershed from another. Finding such typological features based on watersheds will allow a view of where the data does not match the predictions. Anomalies can be identified, which can then be explained through deepening the research model by looking at multiple waves of contact or resettlement, as is found with the Central Bodish languages in eastern Bhutan.

7 Conclusion
This paper applies the basic idea that watersheds are a major factor in delineating language variation in the Tibetic region. At the broad scale, linguistic groupings fall along these geographic lines. At the local level, local language and dialect variation falls along the lines of smaller watersheds.

This is not surprising given the cost-distance effectiveness of travel within a watershed as opposed to across mountain passes. As populations have restricted interaction with those in other watersheds, and have increased interaction with related and unrelated language varieties in their own watershed, language shift and hybridization happen, much like the way that biological systems speciate in these same environments. As in biology, the unique habitat of a watershed, including the geography, flora, fauna, and human populations, affect the languages. The languages are intimately linked to the biology of their watershed.

Through mapping the Tibetic-speaking populations to watersheds, expected patterns of language variation become apparent, as do issues of language maintenance and language shift. From observing what is available based on the geographic analysis, I have laid out a theoretical expectation for what could be found in each region of the Tibetic-speaking area. Some of these can be corroborated with existing data, and others are pure conjecture. Yet even as conjecture, they offer a better understanding of the languages of the area than has previously been available. Now linguistic research can commence in a principled way that will eventually lead to a clearer understanding of the Tibetic region.

Bibliography
(http://resources.arcgis.com/en/help/main/10.1/index.html#/009z000000z5000000)


