Humanities & Social Sciences Communications

ARTICLE

https://doi.org/10.1057/s41599-023-01877-7 OPEN

# Elevation and fog-cloud similarity in Tibeto-Burman languages

Hongdi Ding [1]✉ & Sicong Dong [2]✉

Lexically, 52.99% of the Tibeto-Burman languages, the non-Sinitic branches of the Sino-Tibetan language family, treat fog as something identical or similar to cloud, based on our database of 234 Tibeto-Burman varieties; there are three lexical relations of such fog-cloud similarity in Tibeto-Burman languages, namely cloud colexified with fog, cloud as a hypernym of fog, and cloud as a formative of fog. The rest of the Tibeto-Burman languages use semantically disconnected words to describe fog and cloud. The high proportion of fog-cloud similarity in Tibeto-Burman languages, compared with that of the non-Tibeto-Burman languages spoken alongside the Trans-Himalayan region (i.e., 10.80%, a result based on our database of 213 non-Tibeto-Burman varieties), has its historical reason, namely the relics of Proto-Tibeto-Burman. However, other than the phylogenetic factors, an underlying reason can be attributed to the environmental influence. The present findings indicate that fog-cloud similarity is more likely to happen at higher elevations, particularly between the range of 1000 m to 3000 m above sea level. After reviewing the meteorological features, it is found that the Tibeto-Burman region has ideal conditions for the formation of low cloud, namely with high humidity and through orographic uplift due to the mountainous environment. Since Tibeto-Burman speakers live in high elevations, low cloud, the dominant cloud of the region, may surround them or beneath their view. Therefore, they may find it difficult or not necessary to distinguish fog from low cloud. Our conclusion is also supported by the languages of other families and regions, such as the Daghestanian languages of the Caucasus region and the languages of the Central Andes. Moreover, the present findings agree with the theory of efficient communication. That is, languages displaying fog-cloud similarity are adaptive to higher elevations with less communicative need to distinguish between the two concepts by using completely different and unrelated linguistic forms; on the contrary, languages displaying fog-cloud divergence have stronger need to do so, resulting as well from their adaptation to the extra-linguistic environment. Finally, tropical climates, another possible predictor for fog-cloud similarity, are identified as a future research direction.

[1] The Education University of Hong Kong, Hong Kong, People's Republic of China. [2] Harbin Institute of Technology, Shenzhen, People's Republic of China. ✉email: hding@eduhk.hk; dongsicong@hit.edu.cn

## Introduction

Fog is a cloud resting near the ground; both are aggregates of tiny water droplets or ice crystals suspended in the air (Ahrens, 2012). The difference between fog and cloud is nothing physical but height only. Since clouds are normally high up in the sky and may not disrupt visibility, different from fog that appears near the ground level and can impact daily life, many cultures treat them as different weather events. This is reflected in the use of semantically disconnected words to describe fog and cloud in their languages, such as "cloud" and "fog" in English, and "nuage" and "brouillard" in French.

However, some cultures may experience and perceive fog and cloud as identical or similar weather events. They colexify fog and cloud in their languages, namely, they use the same lexical form for two functionally distinct meanings (François, 2008, p. 170). There are 183 cases of fog-cloud colexification in the database of Cross-Linguistic Colexifications (or CLICS) (Rzymski et al., 2020), such as Blang (Austroasiatic) m̥ut$^{2}$ 'cloud, fog', Lezgian (Nakh-Daghestanian) tsif 'cloud, fog', and Enga (Nuclear Trans New Guinea) mulupána 'cloud, fog'. 123 of the 183 languages, or about 67%, belong to 5 language families in CLICS. One of them is the target family of the present research: the Tibeto-Burman languages[1].

In the present study, we examine the fog and cloud words of the Tibeto-Burman (TB) languages, namely the non-Sinitic branches of the Sino-Tibetan language family (Jacques, 2015). A large number of TB languages, or about 53% in our database of 234 Tibeto-Burman varieties, do not lexically treat fog and cloud as differently as languages like English and French. Some TB languages colexify fog and cloud, such as zdam 'fog, cloud' in Re'ela Qiang (Qiangic) (Zhou, 2019) and t͡ʃa̠m$^{31}$tʰɔi$^{35}$ 'fog, cloud' in Maru (Burmish) (Huang, 1992; Wen, 2022). Some consider fog a hyponym of cloud, such as sazdiə̂m (ground:cloud) 'fog' (cf. zdiə́m 'cloud') in Situ rGyalrong (Qiangic) (Zhang, 2020). In some other TB languages, although fog is expressed with a different morpheme, cloud must be a formative of the fog expression, e.g., də$^{L}$ɹɥɛ̃$^{H}$ (cloud:fog) 'fog' in Niuwozi Prinmi (Qiangic) (Ding, 2014). The three relations are called in this study fog-cloud similarity (cf. fog-cloud divergence in section "Data classification").

Admittedly, there is a phylogenetic reason for fog-cloud similarity in TB languages since they evolved from the common ancestral Proto-Tibeto-Burman (PTB). For example, the fog and cloud words in the above-mentioned Re'ela Qiang, Situ rGyalrong, and Niuwozi Prinmi all retain Proto-Tibeto-Burman *s-dim 'cloud, fog' (Matisoff, 2003). But this leads to the query: why did Tibeto-Burman languages start to exhibit fog-cloud similarity even at the early stages?

Moreover, fog and cloud words in TB languages have multiple etymons. In our database, the fog and cloud expressions can at least be encoded by and traced to eight reconstructed PTB words by Matisoff (2003). Other than *s-dim, the other seven are *r-məw 'sky, heavens, clouds', *muːŋ/*r/s-muːk 'foggy, dark, sullen, menacing, thunder', *kəw-n/t 'smoke', *bʷar/*pʷar 'fire', *m-ka-n 'heavens, sky, sun', *mway 'cloud, fog', and *siŋ/*sik 'wood, firewood, tree'. Similar reconstructions are also found in other sources, such as Benedict (1972), Bradley (1979), Coblin (1986), LaPolla (1987), and VanBik (2009).

However, *bʷar/*pʷar and *mway are not found in cases of fog-cloud similarity, namely not acting as a shared morpheme, which encodes cloud and fog in our TB database. Their reflexes can refer either to fog or cloud, but not both. For example, PTB *bʷar/*pʷar 'fire'[2] is the proto-form of the italicized morpheme in Jingpho (Brahmaputran) sai$^{33}$*wan*$^{31}$ 'fog', with a semantic change from 'fire' to 'fog' (see Burling, 1983; So-Hartmann, 1988), but not used in the cloud words in our database. PTB *mway 'cloud, fog' is used in either cloud words or fog words, but not both, of mainly the Kuki-Chin-Naga languages, such as Tiddim mei$^{2}$ 'cloud', Khumi tmáay 'fog', and Hakha mǐn-*mây* 'cloud' (VanBik, 2009). Although both share the reconstructed meaning of 'cloud', the exact relation between PTB *mway 'cloud, fog'[3] and *r-məw 'sky, heavens, clouds'[4] remains unclear. However, while *r-məw is mainly found as a formative of cloud and fog words in Burmo-Qiangic, Macro-Tani, and Himalayish languages, *mway 'cloud, fog' is mainly used in Kuki-Chin-Naga languages.

Besides *s-dim, mainly found in Burmo-Qiangic languages[5], the other five common etymons (italicized in examples) which are involved in fog-cloud similarity are *r-məw 'sky, heavens, clouds', e.g., doŋ*muk* 'fog, cloud' in Bokar (Macro-Tani) (Huang, 1992; Sun, 1993), *muːŋ/*r/s-muːk 'foggy, dark, sullen, menacing, thunder', e.g., *muk*˥pa˥ 'fog, cloud' in Cangluo Monpa (Bodic) (Zhang, 1986; CASS, 1991), *kəw-n/t 'smoke', e.g., mi$^{55}$*kʰə̰*$^{31}$ 'smoke, cloud, fog' in Yangliu Lalo (Burmo-Qiangic) (Yang, 2010), *m-ka-n 'heavens, sky, sun', e.g., zdeʔm 'cloud' and zdeʔm.*caʔ* (cloud:sky) 'fog' in Kyom-kyo rGyalrong (Burmo-Qiangic) (Prins, 2016; Nagano and Prins, 2013), and *siŋ/*sik 'wood, firewood, tree', e.g., tɕɯ˧ 'cloud' and tɕɯ˧*sɯ*˧˥ 'fog' in Yongning Na (Burmo-Qiangic) (Michaud, 2018). PTB *r-məw and *muːŋ/*r/s-muːk should share a common etymological origin, or have an allofamic relationship, but have developed to the modern languages through different routes (see Matisoff, 2003; Benedict, 1972; LaPolla, 1987).

Therefore, here comes the second query: why do multiple etymons in TB languages, even though 'cloud' and 'fog' may be the derived meanings from the reconstructed meanings (e.g., 'sky', 'smoke', and 'firewood'), end up exhibiting fog-cloud similarity?

The present study aims to seek the underlying reason and answer the following research question: what predicts fog-cloud similarity in Tibeto-Burman languages, other than the phylogenetic relation? The hypothesis is that languages spoken at higher elevations are more likely to exhibit fog-cloud similarity. We will also use the findings to explain the colexification of the non-Tibeto-Burman data in CLICS.

## Literature review

The present study joins the discussion of the influence of the natural environment upon linguistic expressions, which has been a prolific subject of study in the last three decades. There are two major forces in the literature to support linguistic adaptation to ecological conditions. The primary force is the study of the phonetic and phonological patterns (e.g., Munroe et al., 1996, 2009; Munroe and Silander, 1999; Fought et al., 2004; Ember and Ember, 2010; Maddieson, 2012, 2018; Maddieson and Coupé, 2015; Coupé and Maddieson, 2016; Everett et al., 2015; Everett, 2017). Notwithstanding the less impact, perhaps due to smaller sample sizes or less sophisticated algorithms, the lexicon is another main linguistic subsystem, which posits such a relationship with the natural environment (e.g., Witkowski and Brown, 1985; Levinson, 2003; Levinson and Wilkins, 2006; Burenhult and Levinson, 2008; Baddeley and Attewell, 2009; O'Meara and Pérez-Báez, 2011; Palmer, 2015). Discussion from the structural perspective was occasional, e.g., Nichols (1992), although the studies of the influence of other extra-linguistic factors on grammatical structures have been continuous, such as the cultural factors and social factors (e.g., Dunn et al., 2011; see a review in De Busser, 2015).

The lexical perspective, as the theme of the present study, is not new in itself and can be found as early as in Boas's (1911) observation about the words for snow in Eskimo languages (see follow-up discussion in Martin, 1986 and Pullum, 1991) and Sapir's (1912) indication of the "stamps" of the physical environment borne by the vocabulary of a language. With the development of diverse linguistic databases, such as The World Loanword Database (WOLD) (Haspelmath and Tadmor, 2009) and Intercontinental Dictionary Series (IDS) (Key and Comrie, 2015), and the availability of more library references, the environmental impact on the lexicon has gained more attention. For example, Regier et al. (2016) revisited the snow and ice words in the languages of the world and found that languages, which colexify snow and ice tend to be spoken in warmer climates. In other words, people in warmer climates have lower communicative need to distinguish snow and ice. Recently, a series of interdisciplinary studies have looked into the use of verbs in weather expressions (Dong et al., 2020, 2021; Huang et al., 2021). A hypothesis has been proposed by such studies that weather events with bigger weather substances and faster weather processes tend to select action verbs of high transitivity. It has successfully accounted for the selection of verbs in Sinitic weather expressions, e.g., frost is more inclined to use transitive verbs than fog, which is lighter than frost, and the wind expressions using verbs meaning 'to hit' all describe strong wind such as typhoon, which moves much faster than ordinary wind.

Concerning the present hypothesis that languages spoken at higher elevations are more likely to exhibit fog-cloud similarity, two works by Urban (2012, 2023) have also addressed the similar relationship between elevation and the lexical use of fog and cloud, by analyzing the global dataset of IDS and a self-assembled dataset of South American languages. His general finding is that the mean elevation of the languages colexifying fog and cloud is higher than that of the non-colexifying languages (Urban, 2012, 2023). The present study investigates this correlation using different data and methods. Firstly, while Urban (2012, 2023) examined the phenomenon with a focus on the languages of the Central Andes in South America, the present study utilizes data from the Trans-Himalayan region in Asia. The Central Andes feature high elevations and the tropical climate of the Amazon rainforest ecoregions, and both of these environmental variables can affect the lexical use of fog and cloud (see section "Application to CLICS data"). The Trans-Himalayan region, on the other hand, does not feature the tropical climate and we can better observe the impact of elevation.

Secondly, the Tibeto-Burman languages in the present study, or the non-Sinitic branch of the Sino-Tibetan family (or the Trans-Himalayan family), were estimated to be formed around 6000 BP or even earlier, followed by migration and expansion covering topographically and climatically diverse areas (Domrös and Peng, 1988; Shi, 2018; Zhang et al., 2019; Sagart et al., 2019). This time depth is much longer than the languages in the Central Andes, such as Quechuan and Aymaran, which may have evolved around two millennia (Urban, 2023). Therefore, with a longer phylogeny, the Tibeto-Burman languages may have adapted to the environment more effectively, thus allowing us to examine the correlation between the environment and language with higher certainty.

Lastly, Urban's (2012, 2023) findings are based on the "strict colexification" of fog and cloud, namely the exactly same lexeme in synchrony (François, 2008, p. 171), such as gõy 'cloud, fog (as well as smoke)' in Maxakalí, a Nuclear-Macro-Je language in Brazil (Popovich and Popovich, 2005). Differently, the present study samples the data based on both "strict colexification" and "loose colexification" (François, 2008, p. 171), including not only the same lexeme in synchrony but also lexemes which share etymologically related form or exhibit derivational/compounding relationships, such as sazdiə̂m (ground:cloud) 'fog' and zdiə́m 'cloud' in Situ rGyalrong (Qiangic) (Zhang, 2020). By doing so, we can further ground our study into the theory of efficient communication and similar theorizing (Gabelentz, 1901; Bates and MacWhinney, 1982; Du Bois, 1985; Rosch, 1999; Croft, 2003; Haiman, 2010; Regier et al., 2015, 2016). According to Regier et al. (2015, 2016), to support efficient communication, the semantic systems in world languages tend to achieve a near-optimal tradeoff between informativeness and simplicity. The former supports precise communication and the latter minimizes cognitive effort. If a language fulfils its communicative need by strictly colexifying two senses, or "strict colexification", the cognitive effort is the least. However, different languages employ different solutions, which are rated as efficient (Regier et al., 2015). "Loose colexification", like "strict colexification", is also a potential means of minimizing cognitive load, e.g., sharing related forms makes communication cognitively easier than using completely unrelated distinguishing lexemes (see Finley, 2018; Xu et al., 2020), such as 'cloud' and 'fog' in English.

## About Tibeto-Burman languages

Whether Tibeto-Burman is a proper subgrouping under Sino-Tibetan/Trans-Himalayan hypothesis is still controversial (e.g., van Driem, 2007; Jacques and Michaud, 2011). Therefore, we do not use *Tibeto-Burman* in the present study in a subgrouping sense, but only as a term to refer to non-Chinese Sino-Tibetan languages (Jacques, 2015).

The Tibeto-Burman languages comprise about 475 languages spoken across a wide geographic range, or the Tibeto-Himalayan region, mainly in the Hengduan Mountains of southwest China, the Qinghai-Tibet plateau, the Yunnan-Guizhou plateau, Myanmar (formerly Burma), and countries in or beyond the Himalaya, such as Bangladesh, India, Bhutan, Nepal, and Pakistan. The Tibeto-Himalayan region is high in elevation. For example, the average elevation of the Qinghai-Tibet plateau is around 4000 m above sea level; topographically, the Hengduan Mountains, which are to the southeast of the Qinghai-Tibet Plateau, are among the most rugged mountains of the world (Muellner-Riehl, 2019). Due to the ruggedness, biodiversity is promoted, as well as cultural and linguistic diversity (Gorenflo et al., 2012; Axelsen and Manrubia, 2014). Hammarström et al. (2022) classify the TB languages into 17 branches, except the extinct Nam language. The largest three branches are Burmo-Qiangic (158 languages), Kuki-Chin-Naga (87 languages), and Bodic (82 languages). More than half of the 17 branches have only 1 to 3 languages, such as Gongduk (1), Digarish (2), and Kman-Meyor (2).

Moreover, Tibeto-Burman languages have a history of about 6000 years, whose speakers migrated south from the upper reaches of the Yellow River valley into the eastern edge of the Qinghai-Tibet plateau, according to the estimation of the Sino-Tibetan split at the time of the Yangshao Neolithic culture (Zhang et al., 2019). Zhang et al. (2019) also estimate that the initial Tibeto-Burman divergence time, i.e., 4665 years BP, occurred in the middle period of the Majiayao culture, which derived from the Yangshao culture, in eastern Gansu, eastern Qinghai, and northern Sichuan, China. Evidence can still be found in the traditional folklore of the Tibeto-Burman language speakers. For example, speakers of Central Prinmi in Yunnan, a Qiangic language in southwestern China, believe that they are not indigenous to Yunnan, but were originated from an area bordering Qinghai and Gansu to the north of their current home; they also believe that their ancestors led a nomadic life and traveled south until they reached the present-day region between southwestern Sichuan and northwestern Yunnan (Yan and Wong, 1988; Ding, 2014).

**Table 1 A grammatical comparison of selected TB languages.**

| | Manipuri | Tibetic | | Qiangic | | Naic | Lolo-Burmese |
|---|---|---|---|---|---|---|---|
| | Meithei | Lhasa Tibetan | Amdo Tibetan | Yadu Qiang | Prinmi | Naxi | Lahu |
| 1. Copula unused for linking adjectives | - | - | - | + | + | + | + |
| 2. Syllable-tone system | - | - | / | / | - | + | + |
| 3. Lack of verb conjugation | - | - | - | - | +/- | + | + |
| 4. Demonstrative before head noun | - | - | - | - | + | + | +/- |
| 5. Use of classifier when counting people | - | - | - | + | + | + | + |

Tibeto-Burman languages are typologically diverse, containing both isolating languages (e.g., Lolo-Burmese languages) and synthetic languages (e.g., rGyalrongic and Kiranti languages). All TB languages are SOV except the Karenic and Baic branches which are SVO. Most TB languages place modifiers after the noun, although preposed modifiers can also be found (Dryer, 2008). Matisoff (1990, 2003) considers the highly tonal, monosyllabic, and analytic TB languages as the result of Sinospheric influence, and the marginally tonal or atonal TB languages with complex systems of verbal agreement morphology as the result of Indospheric influence. While some TB languages are in one or the other, others have been influenced by both Chinese and Indian cultures. The linguistic features in Table 1 show that while Meithei and Tibetan are more Indospheric, Naxi and Lahu are more Sinospheric; Qiang and Prinmi show mixed features of both.

## Data collection

The fog words and cloud words were collected from 234 Tibeto-Burman languages or dialects from China, Bhutan, Bangladesh, Myanmar, Nepal, and India. They cover 11 branches of the TB languages: Burmo-Qiangic (142), Bodic (33), Kuki-Chin-Naga (16), Himalayish (11), Brahmaputran (11), Macro-Bai (5), Macro-Tani (5), Nungish (4), Kho-Bwa (3), Digarish (2), Dhimalish (1), and Kman-Meyor (1). The sources of data are mainly descriptive grammars, print dictionaries, and three databases: The Sino-Tibetan Etymological Dictionary and Thesaurus (or STEDT[6]), rGyalrongic Languages Database[7], and The Data Collection, Recording, and Display Platform for the Chinese Language Resources Protection Project (or DCRDCLR[8]).

As basic words, expressions for fog and cloud are widely recorded in the sources and their morphological structures can often be clearly analyzed based on the information provided by the sources. We examined all the instances of the fog and cloud words in each source, including the word list and, if available, their usage in phrases and clauses, before we input the form and meaning in our database. We also consulted the relevant part of the reference grammars to understand the morphology of the words when necessary. All the words were cross-checked, wherever possible, by another source(s) of the same variety (e.g., different print references, and the audio files and annotations in DCRDCLR). Typologically, the data can also be cross-checked by the forms of words with the same meaning in varieties of the same language branch. All the data were double-checked after collection (see "Data availability").

For the purpose of comparison, the fog words and cloud words from another 213 languages or dialects were also collected. They are the non-Tibeto-Burman languages, spoken alongside the Trans-Himalayan region which, as defined by Jacques (forthcoming), is a vast area from Baltistan in the West to the Shandong peninsula in the East, and Inner Mongolia in the North down to Myanmar in the South. The comparative languages are spoken at diverse elevations, from as low as 1 m, such as Shenzhen Hakka (Sinitic) in Guangdong, China, to as high as over 3000 m, such as Tajik (Indo-European) in Xinjiang, China. Moreover, the comparative languages represent a high level of linguistic diversity, with a multitude of discrete languages from varied phylogenetic families, covering synthetic (e.g., Indo-European and Turkic) and analytic (e.g., Hmong-Mien) varieties, similar to the TB sample languages. Lexical data from 10 language families were collected (see Fig. 1): Austroasiatic (15), Austronesian (8), Dravidian (4), Hmong-mien (26), Indo-European (12), Mongolic-Khitan (13), Sinitic (72), Tai-Kaidai (42), Tungusic (7), and Turkic (14). The data were also mainly taken from descriptive grammars, print and online databases/dictionaries (e.g., DCRDCLR and Austronesian Basic Vocabulary Database[9]).

To extract the elevation data, we first identified the fieldwork sites or dialectal localities of the data from the references. Then the addresses were searched in Google Earth. To improve accuracy, we recorded the elevations of the data points within 100 m in Google Earth. We also used the coordinates in Glottolog and CLICS, if we cannot identify the exact dialectal localities in the references.

We also extracted the data of annual relative humidity (RH) from Wikipedia when they are available, since an important condition of cloud formation is water vapor or moist air (Ahrens, 2012). Relative humidity is measured by "the ratio of the amount of water vapor in the air to the maximum amount of water vapor required for saturation" (Ahrens, 2012, p. 87). There are 336 RH data obtained out of the 447 sample languages, specifically 162 RH data in the Tibeto-Burman languages and 174 in the comparative languages.

## Data classification

It is oversimplified to treat fog and cloud as different words by merely looking at their lexical forms. While it is easy to make decision about the fog words and the cloud words from 1 to 6 in Table 2 since they are identical, accounting for 32.48% of our TB data, and those from 7 to 12 since they are completely different, accounting for 47.01% of our TB data, morphological and etymological analysis is needed to classify the data such as from 13 to 18, accounting for 20.51% of our TB data. The fog words and the cloud words share a morpheme from 13 to 18 in Table 2. Most of the shared morphemes in Table 2 are reflexes of PTB *s-dim 'cloud, fog' (Matisoff, 2003), such as rGyalrong (Situ) *zdiə́m* 'cloud' and sa*zdiə̂m* 'fog', and Prinmi (Niuwozi) *dĩ*$^{H}$ 'cloud' and *də*$^{L}$ɹɥɛ̃$^{H}$ 'fog'.

Additionally, it is possible for a language to use more than one word for either cloud or fog. Therefore, our classificatory criterion is: a language displays fog-cloud similarity as long as it can express 'fog' and 'cloud' with identical forms or its fog and cloud expressions share the morpheme, which encodes the fog or cloud event. This criterion spares us from being distracted by any complex lexical system for cloud and fog in a particular language. For example, Sherpa (Bodic) distinguishes between shrīn 'high cloud' and mūkpa 'low cloud'. And Sherpa is a case of fog-cloud similarity since mūkpa colexifies 'fog' and 'low cloud' (Hale, 1973; Tournadre et al., 2009). Lahu (Lolo-Burmese) is another example.

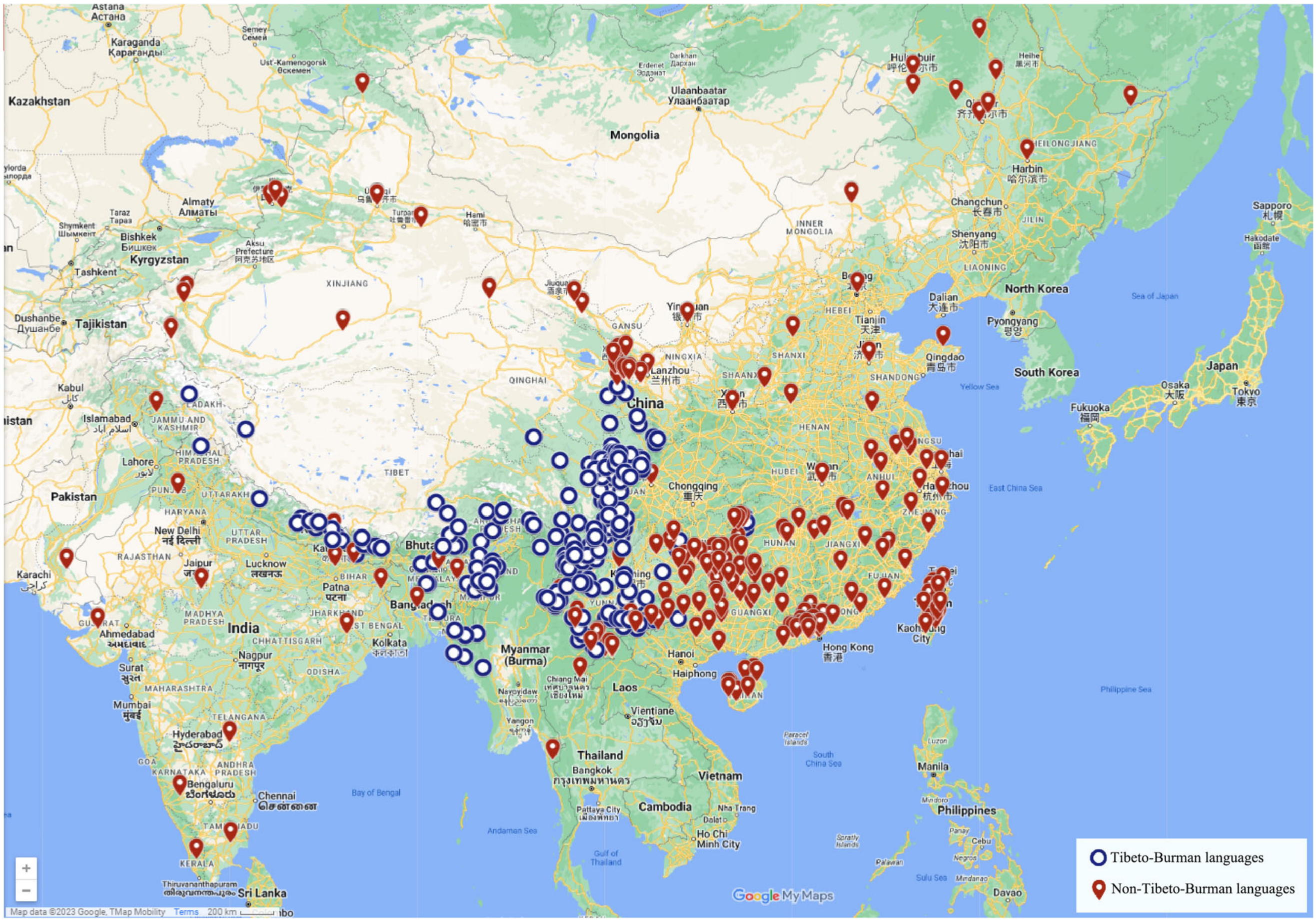



**Fig. 1 Distribution of the sample languages and varieties.** The Tibeto-Burman varieties are concentrated in southwest China and the neighbouring areas of Bhutan, Bangladesh, Myanmar, Nepal, and India. The non-Tibeto-Burman varieties for comparison are distributed alongside the Trans-Himalayan or Sino-Tibetan region.

**Table 2 Fog and cloud words in different Tibeto-Burman languages.**

| | Language | Branch | Cloud | Fog | References |
|---|---|---|---|---|---|
| 1 | Angami Naga (Khonoma) | KCN | kemhu | kemhu | (1) |
| 2 | Bokar | MT | doŋmuk | doŋmuk | (2) |
| 3 | Dumi | H | kɨˈhɨm | kɨˈhɨm | (3) |
| 4 | rGyalrong (Japhug) | BQ | zdɯm | zdɯm | (4) |
| 5 | Manang (Prakaa) | Bodic | $^{3}$mukpə | $^{3}$mukpə | (5) |
| 6 | Qiang (Re'ela) | BQ | zdam | zdam | (6) |
| 7 | Amdo Tibetan (Maqu) | Bodic | htʂən | rm̥ək kwa | (7) |
| 8 | Kokborok | Brh | cumŵi | siyari | (8) |
| 9 | Burmese (spoken) | BQ | tẽ$^{22}$ | mju$^{22}$ | (9) |
| 10 | Gurung (Ghachok) | Bodic | nãː hmjo | ɖuhɖ | (10) |
| 11 | rGyalrong (Geshiza) | BQ | məˈɹŋə | sto-mə | (4) |
| 12 | Zaiwa (Xishan) | BQ | mu̠t$^{55}$mau$^{55}$ | sai$^{55}$von$^{21}$ | (9) |
| 13 | rGyalrong (Situ) | BQ | zdiə́m | sazdiə̂m | (11) |
| 14 | Prinmi (Niuwozi) | BQ | dĩ$^{H}$ | də$^{L}$ɹɥɛ̃$^{H}$ | (12) |
| 15 | Bai (Jianchuan) | MB | vã$^{42}$ | vã$^{42}$kõ̠$^{21}$ | (9) |
| 16 | Qiang (Ronghong) | BQ | zdɑm | zdə-q$^{h}$u | (13) |
| 17 | Khroskyabs (Wobzi) | BQ | zdə̂m | zdə́m-ɲa | (14) |
| 18 | rGyalrong (Kyom-kyo) | BQ | zdeʔm | zdeʔm-caʔ | (15) |

*MT* Macro-Tani, *BQ* Burmo-Qiangic, *KCN* Kuki-Chin-Naga, *H* Himalayish, *Brh* Brahmaputran, *MB* Macro-Bai.
References: (1) Marrison, 1967; (2) Ouyang, 1985; (3) van Driem, 1993; (4) Nagano and Prins, 2013; (5) Hoshi, 1984; (6) Zhou, 2019; (7) Zhou, 2004; (8) Jacquesson, 2008; (9) Huang, 1992; (10) Glover, 1972; Hale, 1973; (11) Zhang, 2020; (12) Ding, 2014; (13) LaPolla and Huang, 2003; (14) Lai, 2017; (15) Prins, 2016.

**Table 3 Fog as a kind or a hyponym of cloud.**

| Language | Branch | Cloud | Fog | References |
|---|---|---|---|---|
| **Fog as "ground cloud"** | | | | |
| rGyalrong (Maerkang) | BQ | zdᴇm | sazdᴇm | (1) |
| rGyalrong (Bola) | BQ | zdem$^{24}$ | sa$^{22}$zdem$^{44}$ | (2) |
| rGyalrong (Hongyuan Shuajinsi Selong) | BQ | zdem | sa'zdem | (2) |
| rGyalrong (Benzhen Yingbolo) | BQ | zdem | sa'zdem | (2) |
| rGyalrong (Taiyanghe) | BQ | zdə$^{m}$ | sazdə$^{m}$ | (2) |
| rGyalrong (Zhuokeji Xisuo) | BQ | zte$^{m}$ | sa'zte$^{m}$ | (2) |
| **Fog as "dark/muddy cloud"** | | | | |
| Khroskyabs (Wobzi) | BQ | zdə̂m̥ | zdə́m̥ɲa | (3) |
| rGyalrong (Erguowucun) | BQ | smək̚pe | smək̚nak̚ | (2) |
| rGyalrong (Jiesi) | BQ | zdem | zdem.mn̥ak̚ | (2) |
| rGyalrong (Taipingqiao) | BQ | zdem | zdem.nak̚ | (2) |
| rGyalrong (Niega Jiaju) | BQ | zdoməʔ | zduɲaʔ | (2) |
| Lisu (Ninglang) | BQ | tɛ$^{35}$mu̠$^{33}$ | tɛ$^{35}$mu$^{33}$xua̠$^{33}$ | (4) |
| **Fog as "prefix-cloud"** | | | | |
| rGyalrong (Ganzi Danba Badi Munashan) | BQ | ᶻdim | kə'zdem | (2) |
| rGyalrong (Jinchuan Maerbang) | BQ | zdiʔm | kə'zdaʔm | (2) |
| rGyalrong (Hongyuan Rangkou Jiadang) | BQ | sa'zdem | kə'zdem | (2) |
| rGyalrong (Lixian Jiabi) | BQ | zdim | kə'zdem | (2) |
| **Fog as "cloud-suffix"** | | | | |
| rGyalrong (Bawang) | BQ | zdo | zdo.mo | (2) |
| rGyalrong (Sawajiao) | BQ | zdim | zdom.wo | (2) |
| rGyalrong (Bajiao) | BQ | zdem | dəɾ.mu | (2) |
| rGyalrong (Daofu) | BQ | zdu | do.mu | (5) |
| nDrapa | BQ | ʂti$^{35}$ | ʂti$^{35}$mbə$^{31}$ | (6) |
| **Fog as "V-ing cloud"** | | | | |
| Lahu (Menglang) | BQ | mu$^{2}$ | mu$^{2}$fei$^{1}$ | (7) |
| Qiang (Mawo) | BQ | zdɤm | zdɤ-qu | (8) |
| Qiang (Ronghong) | BQ | zdɑm | zdə-q$^{h}$u | (9) |

*BQ* Burmo-Qiangic.
References: (1) Huang, 1992; (2) Nagano and Prins, 2013; (3) Lai, 2017; (4) Li, 2022a; (5) Genga, 2019; (6) Gong, 2007; (7) Chang, 1986; (8) Liu, 1998; (9) LaPolla and Huang, 2003.

Although it has various lexical expressions for different types of cloud and fog, as long as we know that 'cloud' and 'fog' can be expressed identically as mò (Matisoff, 2006), it can be concluded that Lahu displays fog-cloud similarity, or specifically a case of fog-cloud colexification. Guiyang[10] Mandarin has two words for 'fog', namely in$^{31}$u$^{24}$ (cloud:fog) 'fog' and u$^{24}$tsau$^{24}$ (fog:covering) 'fog' (Wang, 1994). In Guiyang Mandarin, the fog word in$^{31}$u$^{24}$ (cloud:fog) contains the cloud morpheme in$^{31}$ 'cloud', though the other fog word u$^{24}$tsau$^{24}$ (fog:covering) does not. Since the morpheme which encodes the cloud event is shared by the fog and cloud words, the language is also treated as a case of fog-cloud similarity. Spoken at an elevation of 1274 m, Guiyang Mandarin is the only Sinitic variety of fog-cloud similarity in our database (see section "Higher elevation and fog-cloud similarity").

It is relatively easier to categorize the fog and cloud data as being identical forms and completely different forms. Our focus of the following subsections is on the further sub-categorization of the morpheme-sharing cases. Most of these languages are Burmo-Qiangic, and some are Bodic and Macro-Bai. We have found two major structural relations among them: (1) the cloud morpheme is the head of the fog word, and the other morphemes are modifiers. In this case, fog is understood as a kind or a hyponym of cloud, such as Situ rGyalrong zdiə́m 'cloud' and sazdiə̂m 'fog or ground cloud'; and (2) the cloud morpheme is not the head of the fog word, and it may be a modifier of the fog morpheme or its coordinate. In this case, fog is not a kind or a hyponym of cloud, such as dĩ$^{H}$ 'cloud' and də$^{L}$ɹɥɛ̃$^{H}$ (cloud:fog) 'fog' in Niuwozi Prinmi, and in$^{31}$ 'cloud' and in$^{31}$u$^{24}$ (cloud:fog) 'fog' in Guiyang Mandarin. It is also discovered most Tibeto-Burman languages use more complex morphological structures for fog, often based on the cloud morphemes. The word formations of the fog words are through derivation and compounding (modification and coordination). Some cases can be found where the cloud word is based on the fog morpheme. In Yangliu Lalo and Mangdi Lalo, both Lolo-Burmese varieties under the Burmo-Qiangic branch, the cloud words, namely mi$^{55}$k$^{h}$ɔ̪$^{31}$ and mi$^{55}$kɨ$^{21}$ respectively[11], are formed based on 'fog' mi$^{55}$ and 'smoke' k$^{h}$ɔ̪$^{31}$/kɨ$^{21}$ (Yang, 2010).

**Fog is a kind or a hyponym of cloud**. When the cloud morpheme is the head of the fog word in the word formation, fog is understood as a hyponym of cloud.

*Fog is "ground cloud"*. Cross-linguistically, it is common for fog to be called literally as "ground cloud", such as Bonan (Mongolic-Khitan) ɢɑdʐir mokə (ground cloud) 'fog' (Ding, 2022) and Pnar (Austroasiatic) lʔɔʔ k$^{h}$n̥daw (cloud ground) 'fog'[12] (Nagaraja et al., 2013). In our Tibeto-Burman data, as is exemplified by rGyalrong (Situ) in Table 2, the fog word is compounded with two nominal formatives: sa and zdiə́m. The former is a reflex of PTB *(s/z)a-y 'earth, ground, soil, sand' and the latter PTB *s-dim 'cloud, fog' (Matisoff, 2003), hence literally "ground cloud" (see Table 3).

*Fog is "dark/muddy cloud"*. As is exemplified by Khroskyabs (Wobzi) in Table 2, the fog word is compounded through the cloud morpheme and a postposed morpheme meaning 'dark or black, muddy', hence literally meaning 'dark cloud' or 'muddy cloud' since most Tibeto-Burman languages place the modifier of property after the head noun. This pattern is also found in Qiangic, such as dámə̰ (cloud:dark) 'fog, cloud' in Longxi Qiang and dámò (cloud:dark) in Mianchi Qiang (Evans, 1999; Zheng, 2016), and rGyalrongic languages (see Table 3), and Lolo-Burmese languages (e.g., Ninglang Lisu) (see Table 3). The rGyalrongic modifying morphemes mean 'dark, black', all of them being reflexes of Proto-Tibeto-Burman *s-ma(ŋ/k) / *s-nak 'ink, black, deep', reconstructed by LaPolla (1987) and Matisoff (2003). Lisu morpheme xua̰$^{33}$ means 'muddy' (Li, 2022a); but its source is not clear.

*Fog is "prefix-cloud"*. Again, in rGyalrongic languages, the prefix kə- is probably historically related to the velar nominalization prefix, reconstructed as *gV-. See a cross-linguistic discussion of the PTB prefix *gV- in Konnerth (2016). Its functions in rGyalrongic languages, as well as other TB branches (e.g., Kuki-Chin-Naga and Brahmaputran), include derivational nominalization and clausal nominalization (see Sun, 2014; Nagano, 2017; Jacques, 2021). Specifically, the prefix kə- should create gerund nominalization for the fog expression of the rGyalrongic varieties in Table 3, literally meaning 'being cloudy'.

*Fog is "cloud-suffix"*. There are two major types of suffixes in our TB data, namely the reflexes of PTB nominalizer *-pu / *-pwa and of PTB gender suffixes (Benedict, 1972; Matisoff, 2003). It is a common derivation in Bodic languages to express 'cloud' and 'fog' with the nominalizer (italicized), such as mu:*pa* 'fog' in Kaike (Hale, 1973) and tʂĩ$^{55}$*pə$^{55}$* 'cloud' in Lhasa Tibetan (Huang, 1992). In our TB data, the suffix -mbə$^{31}$ in nDrapa (Burmo-Qiangic) ʂti$^{35}$*mbə$^{31}$* (cloud-nominalizer) 'fog' should be a borrowing from the Tibetic language (Huang, 2020). Since the stem ʂti$^{35}$ of the fog word is a reflex of PTB *s-dim 'cloud, fog', the core meaning of the derived word is not changed. Regarding the gender suffix, Honkasalo (2019: p. 225) points out that Eastern Geshiza rGyalrong zdo-ma 'cloud' borrows the suffix *-ma* from Tibetan, related to the historical feminine suffix (also see Matisoff, 1991). The rGyalrong suffixes *-mo*/*-mu*/*-wo* in Table 3 should all be the gender suffixes. While *-mo*/*-mu*, similar to Eastern Geshiza *-ma*, are probably based on the Tibetan feminine nominal suffix *-mo*, *-wo* is from the Tibetan masculine nominal suffix *-po*.

*Fog is "V-ing cloud"*. This formation involves the use of the cloud formative and a verbal formative. In Menglang Lahu (Lolo-Burmese), the morpheme fei$^{1}$ in the fog word mu$^{2}$fei$^{1}$ means 'to cover something up', semantically similar to the verb fiʔ in Black Lahu (Matisoff, 2006). Therefore, literally, fog in Menglang Lahu means 'covering cloud'. This kind of N-V compounding is also found in Qiangic languages. For example, in Ronghong Qiang, zdə.q$^{h}$u (cloud:descend) refers to 'fog' and zdɑm to 'cloud' (LaPolla and Huang, 2003); similarly, in Mawo Qiang, zdɤ.qu (cloud:descend) means 'fog' and zdɤm 'cloud' (Liu, 1998). Therefore, in Ronghong and Mawo Qiang, the meaning of 'fog' is literally "descending cloud". Nouns formed via N-V compounding are popular in TB languages, such as me$^{ɹ}$gu̥ 'thunder' < meː$^{ɹ}$ 'sky' + gu 'to thunder' in Ronghong Qiang (LaPolla and Huang, 2003, p. 332).

*Unidentified modifying morpheme*. It is sometimes unable to identify the origins of some modifying morphemes, but decision can still be made about their sub-categorization. For example, the source of the morpheme ʐø$^{35}$ in Shade Muya (Burmo-Qiangic) ʐø$^{35}$ndɯ$^{33}$ʐe$^{35}$ 'fog' is unknown, where ndɯ$^{33}$ʐe$^{35}$ refers to 'cloud' (CASS, 1991); bo$^{33}$ in Ersu (Burmo-Qiangic) bo$^{33}$tsɛ$^{55}$ 'fog' is unclear about its source, where tsɛ$^{55}$ refers to 'cloud' (CASS, 1991). Since the morpheme preceding the cloud word is not found to be a coordinate, but either a nominal modifier or a prefix in our sample TB languages, the cloud morpheme is highly likely to be the head of the compounding and fog is as well a kind of cloud. It is suspected that ʐø$^{35}$ in Shade Muya and bo$^{33}$ in Ersu are both loanwords from Southwest Mandarin, namely ʐø$^{35}$ is related to Southwest Mandarin jy$^{53}$ 'rain' and bo$^{33}$ Southwest Mandarin po$^{21}$ 'thin'. Regarding the former, cognitively, it is possible for people to use water-related concepts to refer to fog (see section "Fog is 'cloud water'"). Regarding the latter, when an adnominal modifier is borrowed, it is common for the borrowed Chinese adjective/stative verb to be used before the head noun. For instance, in Liangshan Yi, with which Ersu has frequent contact, the first morpheme ta$^{55}$ of the word ta$^{55}$ga$^{33}$ (big:road) 'big road' is a loanword from Southwest Mandarin ta$^{213}$ 'big', although there is an inherent expression ga$^{21}$mo$^{21}$ (road:big) '(big or main) road' in Liangshan Yi.

**Fog is not cloud, but involves cloud**. Unlike the hyponym-hypernym relation of fog and cloud in section "Fog is a kind or a hyponym of cloud", cloud is not the head morpheme of the word formation, but a modifier or a coordinate component of the fog word. It is also observed that TB languages commonly relate fog to other concepts in these expressions, such as ash, smoke, and dew.

*Fog is "cloud ash"*. In Dechang and Yongsheng Lisu, the second morphemes (italicized in examples) of the fog words, namely mu$^{44}$ and m̩$^{44}$, refer to 'ashes, dust', such as na$^{44}$ts$^{h}$ɿ$^{31}$*mu$^{44}$* (medicine:ash) 'medicine powder' and ʃa$^{44}$*mu$^{44}$* (wheat:ash) 'flour' in Dechang Lisu (Li, 2022b), and na$^{44}$ts$^{h}$ɿ$^{42}$*m̩$^{44}$* (medicine:ash) 'medicine powder' and dza$^{33}$*m̩$^{44}$* (grain:ash) 'flour' in Yongsheng Lisu (Li, 2022c). It is common to find in other languages of the world the colexification of 'ashes, dust' and 'fog'/ 'cloud', such as Wabula Cia-Cia (Austronesian) gaβu 'dust, fog' (Kaiping et al., 2019), Buyang (Tai-Kadai) la$^{0}$muk$^{11}$ 'dust, fog' (Key and Comrie, 2015), and Bukusu (Atlantic-Congo) fuumbi 'dust, cloud' (Greenhill and Gray, 2015). In Tibeto-Burman languages, Burmese (written) mru also displays this kind of colexification, namely 'minute particle; mist, fog' (Benedict, 1976).

This type of compounding is also identified in Naic and Bodic languages but with possible semantic extension. In Naxi and Yongning Na (Narua) (see Table 4), two Naic languages, the first morphemes of the fog words, namely tɕi$^{31}$ and tɕɯ˧, refer to 'cloud'; the second morphemes sɯ$^{33}$ and sɯ˧˥ are reflexes of PTB *si(ŋ/k) 'wood, firewood, tree' or PST *siŋ 'wood, firewood, tree' (Chou, 1972; LaPolla, 1987; Matisoff, 2003). This diachronic relation is also consistently found in synchronic Naic data between 'fog' and 'firewood', such as Dayan Naxi tɕ$^{h}$i$^{55}$sɚ$^{33}$ 'fog' and sɚ$^{33}$ 'firewood' (Zhao, 2022), and Yanbian Naxi tsɿ$^{21}$sɿ$^{33}$ 'fog' and sɿ$^{33}$ 'firewood' (Liu, 2022). There should be a further semantic extension of the second morpheme from 'firewood' to 'ash', probably via an intermediate connection with 'charcoal'[13]. The path of semantic development from 'charcoal' to 'ash' is also typologically attested by Sunwar (Himalayish) koylāː 'charcoal, ash' (Hale, 1973), and Botlikh (Nakh-Daghestanian) кьей 'charcoal, ash' (Key and Comrie, 2015).

*Fog is "cloud smoke"*. In Luquan Lisu (see Table 4), the fog word is formed by the formative ti$^{33}$ 'cloud' and k$^{h}$ə$^{31}$/k$^{h}$e$^{31}$ 'smoke' (Mu and Sun, 2012), where the former is a reflex of PTB *s-dim 'cloud, fog' and the latter a reflex of PTB *kəw-n/t 'smoke'. Therefore, there is a connection between smoke and fog in Luquan Lisu. Some languages colexify fog and smoke, such as Batsbi (Nakh-Daghestanian) k'ur 'fog, smoke' (Carling, 2017) and Rongga (Austronesian) nuː 'fog, smoke' (Kaiping et al., 2019).

*Fog is "cloud dew"*. In Bai, the fog word vã$^{42}$kõ̰$^{21}$ is formed with the formative 'cloud' and 'dew'. Although the fog expression must

**Table 4 Fog is not cloud, but involves cloud.**

| Language | Branch | Cloud | Fog | References |
|---|---|---|---|---|
| **Fog as "cloud ash"** | | | | |
| Naxi (Qinglongxiang) | BQ | tɕi$^{31}$ | tɕ$^{h}$i$^{55}$sɯ$^{33}$ | (1) |
| Yongning Na | BQ | tɕɯ˧ | tɕɯ˧sɯ˧˥ | (2) |
| Lisu (Dechang) | BQ | mu$^{31}$ti$^{33}$ | ti$^{35}$mu$^{44}$ | (3) |
| Lisu (Yongsheng) | BQ | m̩$^{31}$ti$^{44}$ | ti$^{35}$m̩$^{44}$ | (4) |
| **Fog as "cloud smoke"** | | | | |
| Lisu (Luquan) | BQ | ti$^{33}$t͡ʃ$^{ɦ}$o$^{33}$ | ti$^{33}$k$^{h}$e$^{31}$ | (4) |
| **Fog as "cloud dew"** | | | | |
| Bai (Jianchuan) | MB | vã$^{42}$ | vã$^{42}$kõ̠$^{21}$ | (5) |
| **Fog as "cloud sky"** | | | | |
| rGyalrong (Kyom-kyo) | BQ | zdeʔm | zdeʔm-caʔ | (6) |
| rGyalrong (Jinchuan Kalajiao) | BQ | zdim | zde'mk$^{h}$e | (7) |
| rGyalrong (Maerkang Songgang Zhibo) | BQ | zdem | zdem'ca | (7) |
| rGyalrong (Xiaojin Mupo) | BQ | zdem | zdem'k$^{h}$ɑ | (7) |
| rGyalrong (Xiaojin Rilong) | BQ | $^{z}$d$^{i}$ɛm | $^{z}$d$^{i}$ɛm'k$^{h}$aʔ | (7) |
| **Fog as "cloud water"** | | | | |
| Muya (Pengbuxi) | BQ | ndə$^{33}$ʐe$^{53}$ | ndɛ$^{33}$tɕ$^{h}$ʌ$^{53}$ | (8) |
| **Fog as "cloud steam"** | | | | |
| Yi (Lalo) (Shuizhuping) | BQ | ti$^{24}$ | ti$^{24}$kv̩$^{21}$ | (9) |
| **Fog as "cloud and fog"** | | | | |
| Pumi (Dayang) | BQ | zdin$^{55}$ | zdɛ$^{31}$ʐən$^{55}$ | (10) |
| Pumi (Jiulong) | BQ | dẽ$^{35}$ | zɿ$^{55}$ʐẽ$^{55}$ | (5) |
| Pumi (Ludian) | BQ | dĩ$^{55}$ | di$^{13}$rã$^{55}$ | (11) |
| Pumi (Niuwozi) | BQ | dĩ$^{H}$ | də$^{L}$ɹɥɛ̃$^{H}$ | (12) |
| Pumi (Qinghua) | BQ | sdĩ$^{55}$ | sdie$^{13}$ʐə̃$^{55}$ | (13) |
| Qiang (Taoping) | BQ | χde$^{33}$ | χde$^{33}$le$^{33}$ | (14) |
| Yi (Manshuiwan) | BQ | mu$^{33}$vu$^{55}$ | mu$^{33}$vu$^{55}$vu$^{55}$ | (15) |

*MB* Macro-Bai, *BQ* Burmo-Qiangic.
References: (1) He and Jiang, 1985; (2) Michaud, 2018; (3) Li, 2022b; (4) Mu and Sun, 2012; (5) Huang, 1992; (6) Prins, 2016; (7) Nagano and Prins, 2013; (8) Gao, 2022; (9) Yang, 2010; (10) Jiang, 2015; (11) Lu, 2001; (12) Ding, 2014; (13) Lu, 1983; (14) CASS, 1991; (15) Qumu, 2022.

contain the cloud morpheme in Bai, some languages can colexify dew and fog with identical forms, such as Wancho (Brahmaputran) rangphum 'dew, fog' (Marrison, 1967), and Romani (Indo-European) bruma 'dew, fog' (Key and Comrie, 2015).

*Fog is "cloud sky"*. Fog expression in rGyalrongic languages (see Table 4) can also be formed by compounding PTB *s-dim 'cloud, fog' and PTB *m-ka-n 'heavens, sky, sun', such as rGyalrong (Kyom-kyo) zdeʔm.caʔ (cloud:sky) 'fog', rGyalrong (Xiaojin Zhailong) zdem.k$^{h}$ɑ (cloud:sky) 'fog', and rGyalrong (Lixian Ganbao) zəŋ.k$^{h}$e (cloud:sky) 'fog' (Nagano and Prins, 2013). Since both formatives are nominals, the cloud morpheme is not the head of the fog word, but a modifier. Fog thus literally means "cloud sky".

*Fog is "cloud water"*. In Pengbuxi Muya, the fog word *ndɛ*$^{33}$tɕ$^{h}$ʌ$^{53}$ shares the cloud morpheme (italicized) with the cloud word *ndə*$^{33}$ʐe$^{53}$. The other morpheme tɕ$^{h}$ʌ$^{53}$ is a variant of the word tɕʌ$^{53}$ 'water' in Muya, which may become aspirated in compounding, namely ndɛ$^{33}$tɕ$^{h}$ʌ$^{53}$. Associating 'fog' with water is also found in Sinitic languages, such as Liuzhou Mandarin (Sinitic) u$^{24}$suɐi$^{54}$ (fog:water) (Liu, 1995) and Dongguan Yue (Sinitic) mɔu$^{32}$sui$^{35}$ (fog:water) 'fog' (Zhan et al., 1997). This connection also conforms to the physical properties of fog as a form of water (Day, 1998; Ahrens, 2012).

*Fog is "cloud steam"*. In Shuizhuping Lalo, the fog word is compounded with the cloud morpheme ti$^{24}$ and the steam morpheme kv̩$^{21}$ (see Table 4) (Yang, 2010). Colexification of steam and fog is commonly attested in other languages, such as Romanian (Indo-European) abur 'steam, fog', and Otomi (Otomanguean) 'bipa 'fog, steam' (Haspelmath and Tadmor, 2009).

*Fog is "cloud and fog"*. This formation is through coordinate compounding of the cloud morpheme with the fog morpheme, namely 'fog' < cloud + fog, such as Prinmi (Niuwozi) də$^{L}$ɹɥɛ̃$^{H}$. The fog morphemes in our database have diverse etymons. For example, the fog morphemes in the Prinmi[14] varieties and Qiang are probably cognate with le 'fog' in Tangut, the extinct Qiangic language (see Li, 1997 and Table 4). Tangut le is still kept in χde$^{33}$le$^{33}$ (cloud:fog) 'fog' of Taoping Qiang, a southern Qiang dialect.

In Manshuiwan Yi, the fog morpheme vu$^{55}$, probably a Southwest Mandarin loanword, is lexicalized to be part of the cloud word mu$^{33}$vu$^{55}$ (cloud:fog) 'cloud'; the fog word is expressed with an additional fog morpheme mu$^{33}$vu$^{55}$vu$^{55}$ (cloud:fog) 'fog'. In this kind of formation, there is a specific morpheme for fog; and cloud, not being the head of the compounding, is a formative of the fog expression. In other words, cloud may be considered a necessary component of fog in these cultures.

**Summary**. After the morphological analysis, four types of data are identified in the database. For the first type of data, fog is cloud, identically, such as Lizu, tɕe$^{53}$ 'fog, cloud' (Huang, 1992). This type of data displays fog-cloud colexification. For the second type of data, fog is also cloud, but with modifications, acting as cloud's hyponym, such as zdiə́m 'cloud' and sazdiə̂m (ground:cloud) 'fog' in rGyalrong (Situ). For the third type of data, fog is not cloud, but involves the concept of cloud, such as dĩ$^{H}$ 'cloud' and də$^{L}$ɹɥɛ̃$^{H}$ (cloud:fog) 'fog' in Prinmi (Niuwozi). For the last type of data, fog is completely different from and unrelated to cloud, such as ti$^{33}$ 'cloud' and mu$^{33}$n̥o$^{55}$ (sky:fog) 'fog' in Liangshan Yi. The first three types of data are called fog-cloud

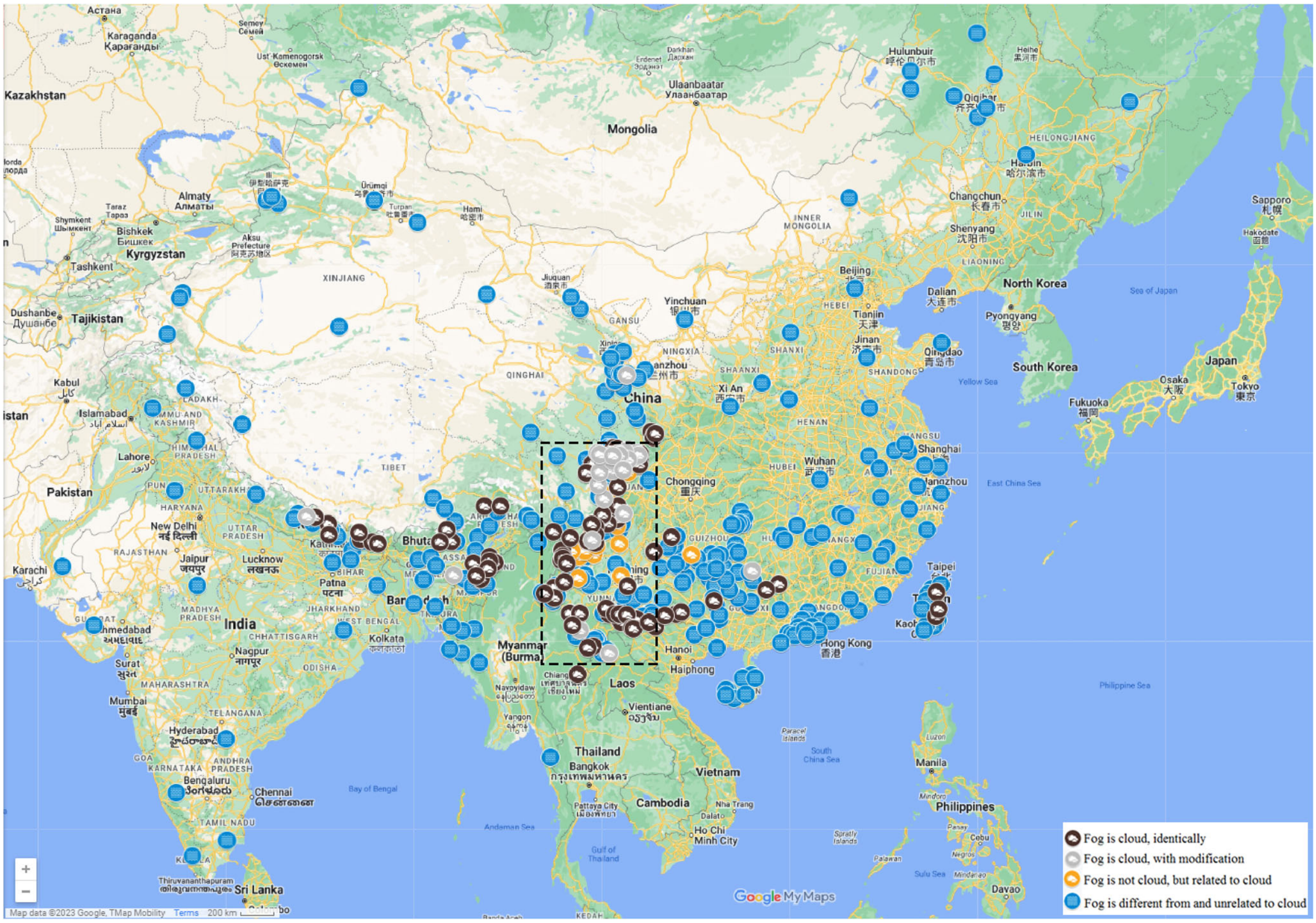



**Fig. 2 Distribution of fog-cloud similarity and fog-cloud divergence of the sample languages.** The languages in the dotted square are to the southeast of the Qinghai-Tibet Plateau, an area which features high cloud cover and high relative humidity.

similarity in the present study, and the fourth type is fog-cloud divergence. We processed the non-Tibeto-Burman data in the same way. See the distribution of fog-cloud similarity and fog-cloud divergence of the sample languages in Fig. 2. Due to the lack of lexical and morphological information, there are five TB data points in our collection, which we cannot further sub-categorize, namely Maram Naga (Kuki-Chin-Naga) *kamong* 'cloud' and *kamong*-sole 'fog' (Marrison, 1967), Puroik (Kho-Bwa) *kə*$^{33}$*tɯ*$^{33}$ and *kə*$^{33}$*tɯ*$^{33}$sɯ$^{33}$ (CASS, 1991), Gyaru Manang (Bodic) *mɯʔ*$^{2}$pa$^{2}$ 'cloud' and *muk*$^{2}$sɯl$^{2}$ 'fog' (Nagano, 1984), Mianning Namuyi (Burmo-Qiangic) *tʂ̺u*$^{33}$ 'cloud' and *tʂ̺u*$^{33}$tɕ̺hi$^{33}$xo$^{35}$ 'fog' (CASS, 1991), and Tuoqi Prinmi *də*$^{13}$rõ$^{53}$ 'cloud' and *də*$^{13}$rẽ$^{55}$ 'fog' (Lu, 2001). Although whether they should be sub-categorized as the second or third type remains undetermined, it is still safe to conclude that these data points show fog-cloud similarity since the cloud morpheme (italicized above) is contained in the fog word. The first two lexical relations, namely fog-cloud colexification and fog as a hyponym of cloud, form the core of fog-cloud similarity since there is no specific word for fog. The third type, i.e., cloud as a formative of fog, can be considered as the transitional layer from core fog-cloud similarity to fog-cloud divergence since there comes a specific morpheme for fog. It is also noted that fog-cloud similarity in Tibeto-Burman languages is mostly concentrated to the southeast of the Qinghai-Tibet plateau (see the dotted square in Fig. 2).

## Results and discussion

In this section, we will discuss the environmental influence, the hypothesized underlying reason besides the phylogenetic relations, for fog-cloud similarity in Tibeto-Burman languages. It is also found that language contact is a major reason for relatively recent fog-cloud similarity and divergence. Finally, we will apply our findings to the colexification data in the database CLICS.

**Higher elevation and fog-cloud similarity**. In our database, fog-cloud similarity accounts for 52.99% of the Tibeto-Burman languages, but only 10.80% of the non-Tibeto-Burman data. The TB and non-TB data also suggest that languages displaying fog-cloud similarity have higher average and median elevations than fog-cloud divergence languages. See Table 5. We ran a Two-Sample t-Test in Excel. The result shows that the elevations of fog-similarity languages are significantly different from those of fog-cloud divergence languages. Similar findings were reported in Urban (2023) by using the IDS and Central Andean data of "strict colexification".

Meanwhile, the range of elevation is also narrower in fog-cloud similarity languages than in fog-cloud divergence languages, suggesting that fog-cloud similarity is least likely to occur in some elevations. The top four ranges of elevation where fog-cloud similarity is found in TB languages are from 1000–1500 m, 1500–2000 m, 2000–2500 m, and 2500–3000 m (see Fig. 3). If the elevation is lower than 500 m or higher than 3500 m, fog-cloud similarity is unlikely to occur. This observation is also valid if only the core fog-cloud similarity TB languages and all the TB and non-TB data are considered. This is a main different discovery from Urban (2023): in his study, colexifying languages were spoken at both low and high elevations; in other words, there are fewer restrictions on the distribution of colexification, which is in

**Table 5 Elevation and fog-cloud similarity/divergence.**

| **Tibeto-Burman languages** | | **Percentage** | **Elevation** | | | |
|---|---|---|---|---|---|---|
| | | | **Average** | **Max** | **Min** | **Median** |
| Type 1: fog is cloud | 76 | 32.48% | 1957.5 | 4357 | 643 | 1779.5 |
| Type 2: fog is a kind of cloud | 22 | 9.4% | 2777.0 | 3698 | 1017 | 2924 |
| Type 3: fog is related to cloud | 21 | 8.97% | 2389.6 | 3407 | 1274 | 2520 |
| Unclassified data points: fog is a kind of or related to cloud | 5 | 4% | | | | |
| TOTAL of fog-cloud similarity | 124 | 52.99% | 2213.5 | 4357 | 643 | 2190 |
| TOTAL of fog-cloud divergence | 110 | 47.01% | 1726.4 | 4693 | 5 | 1687 |
| *P*-value | | | 0.00 (<0.05) | | | |
| **non-Tibeto-Burman languages** | | **Percentage** | **Elevation** | | | |
| | | | **Average** | **Max** | **Min** | **Median** |
| Type 1: fog is cloud | 17 | 7.98% | 1203.1 | 2287 | 235 | 1310 |
| Type 2: fog is a kind of cloud | 4 | 1.88% | | | | |
| Type 3: fog is related to cloud | 2 | 0.94% | | | | |
| TOTAL of fog-cloud similarity | 23 | 10.80% | 1206.3 | 2287 | 235 | 1274 |
| TOTAL of fog-cloud divergence | 190 | 89.20% | 542.9 | 3090 | 1 | 235.5 |
| *P*-value | | | 0.00 (<0.05) | | | |

contradistinction to the findings in Regier et al. (2016). On the contrary, the present study supports Regier et al. (2016). That is, the colexifying languages are more strongly constrained than the diverging languages with regard to the non-linguistic variables, temperature in Regier et al.'s (2016) snow-and-ice case and elevation in the present fog-and-cloud study.

To account for the discrepancy, Urban (2023) ascribed to lineage-specific preferences, namely a language family can be consistently colexifying, such as the Quechuan family, or consistently differentiating, such as the Aymaran family. Our results partly agree with the lineage-specific account: the lineage-specific preference can be observed at the lower end of the family tree. In our samples, the three largest branches of the Tibeto-Burman languages, namely Burmo-Qiangic, Kuki-Chin-Naga, and Bodic, feature both fog-cloud similarity languages and divergence languages, showing little evidence of intra-lineage effect at such higher-level nodes. For example, 35.97% of the Burmo-Qiangic samples distinguish fog and cloud with completely unrelated forms, and 35.25% strictly colexify fog and cloud. Similarly, both strictly colexifying and completely differentiating languages are found in the Bodic branch, with 25% of the former and 71.9% of the latter. Most of the diverging languages within the Bodic branch at very high elevations, above 3000 m, come from the Tibetan varieties, showing the lineage-specific effect at the lower-level node. But the lineage-specific effect may not be at play at other lower-level nodes. For example, in our non-TB samples, both strictly colexifying and completely differentiating languages are found in Miao (Hmongic) and Bouyei (Kam-Tai). Among the 12 Miao varieties, the only two colexifying fog and cloud are located at the elevations of 1431 m and 1722 m, while the other ten differentiating fog and cloud with unrelated forms average 701.1 m, ranging from 351 m to 1086 m. The only colexifying Bouyei has the highest elevation among the three Bouyei varieties in our samples, namely 2107 m versus 1094 m and 1275 m.

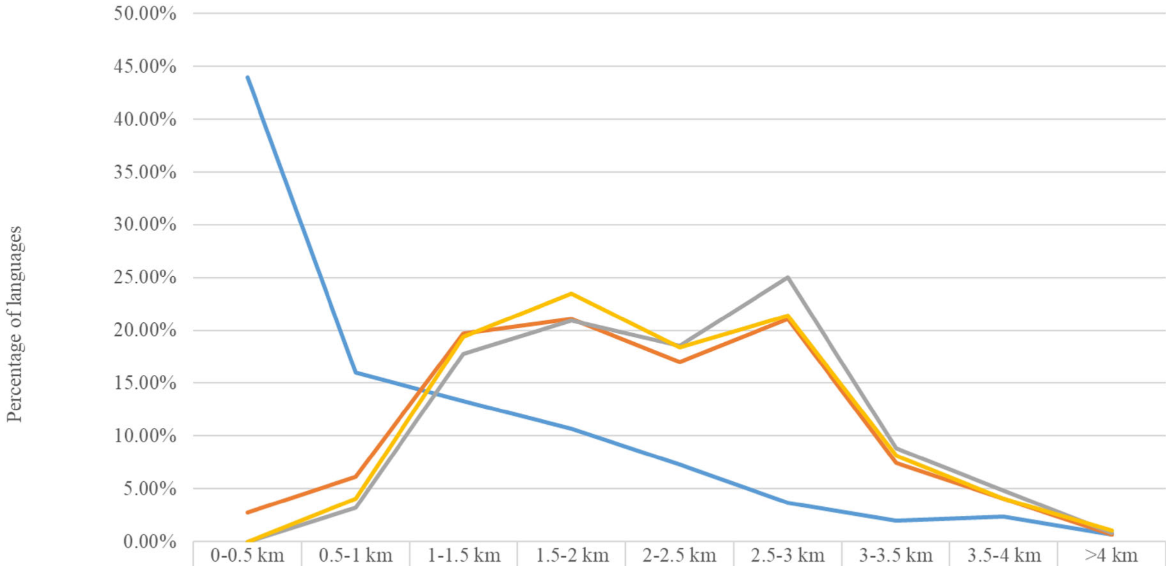
Besides, we examined the locations of the Central Andean colexifying data below 500 m in Urban (2023) and found that all of them fell within the Amazon rainforest ecoregions featuring the tropical climate. Instead of a lineage-specific preference, the colexification of fog and cloud in these languages is probably the result of adaptation to the tropical climate, which is another extra-linguistic variable for this phenomenon (see section "Application to CLICS data").

Additionally, people opt to settle down at lower elevations (Nogués-Bravo et al., 2008), namely, there should be more languages spoken in lower areas. Even given this correlation between settlement distribution and elevation, however, fog-cloud similarity still shows robust relations with higher elevations. In other words, the number of languages of fog-cloud divergence decreases as elevation increases, showing a general settlement tendency; however, the distribution of fog-cloud similarity is not related to the settlement pattern (see Fig. 3).

**A mixture of low cloud and fog**. Fog-cloud similarity is most likely to occur between elevation 1000 m and 3000 m in the Tibeto-Burman area. Two kinds of cloud also occur in this range in the middle-latitude region, or the subtropical and temperate zones (cf. the tropical zone in section "Application to CLICS data"), namely the low cloud (0–2000 m) and midlevel cloud (2000–7000 m) (Ahrens, 2012, p. 103).

Liu et al. (2018) and Wei et al. (2020) indicate that the southeast of the Qinghai-Tibet plateau, the hotspot of fog-cloud similarity (see Fig. 2), is heavily overcast, with annual total cloud cover up to 69.5%, due to the high relative humidity by moisture transport from the Bay of Bengal. The average annual relative humidity of the places where we found fog-cloud similarity is 67.87%, ranging from 42% in Shannan, Tibet, China, to 80% in Lianghe County, Dehong Dai and Jingpo Autonomous Prefecture, Yunnan, China. Moreover, low cloud is the dominant cloud in this area, with an annual low cloud cover of 51.9% (Wei et al., 2020). According to Walcek (1994), cloud cover is positively correlated with the relative humidity of a region. Similarly, a high level of low cloud cover can also be found in the southern slope of the Himalaya due to the monsoon, and the frequency of cloud coverage can exceed 75% at 15 Local Solar Time in the monsoon period (Jaswal et al., 2017; Kattel et al., 2013; Kurosaki and Kimura, 2002). Comparatively, since the west of the Qinghai-Tibet plateau is more arid, it has less cloud cover: its annual total cloud cover and annual low cloud cover are 49% and 30.5%, respectively (Wei et al., 2020).

Liu et al. (2018) also indicate that in the southeast of the Qinghai-Tibet plateau, the most frequent low clouds are stratus




| | 0-0.5 km | 0.5-1 km | 1-1.5 km | 1.5-2 km | 2-2.5 km | 2.5-3 km | 3-3.5 km | 3.5-4 km | >4 km |
|---|---|---|---|---|---|---|---|---|---|
| fog-cloud divergence in all data | 44.00% | 16.00% | 13.33% | 10.67% | 7.33% | 3.67% | 2.00% | 2.33% | 0.67% |
| fog-cloud similarity in all data | 2.72% | 6.12% | 19.73% | 21.09% | 17.01% | 21.09% | 7.48% | 4.08% | 0.68% |
| fog-cloud similarity in TB lgs | 0.00% | 3.23% | 17.74% | 20.97% | 18.55% | 25.00% | 8.87% | 4.84% | 0.81% |
| core fog-cloud similarity in TB lgs | 0.00% | 4.08% | 19.39% | 23.47% | 18.37% | 21.43% | 8.16% | 4.08% | 1.02% |

fog-cloud divergence in all data — fog-cloud similarity in all data — fog-cloud similarity in TB lgs — core fog-cloud similarity in TB lgs

**Fig. 3 Fog-cloud divergence languages, fog-cloud similarity languages, and elevation.** Fog-cloud similarity is least likely to occur if the elevation is below 500 m and above 3500 m. The top four ranges of elevation where fog-cloud similarity is found in TB languages are from 1000–1500 m, 1500–2000 m, 2000–2500 m, and 2500–3000 m. Moreover, languages of fog-cloud divergence decrease as elevation increases, showing a general tendency that people prefer settling down in areas of lower elevations. However, the distribution of fog-cloud similarity is not related to the settlement distribution.

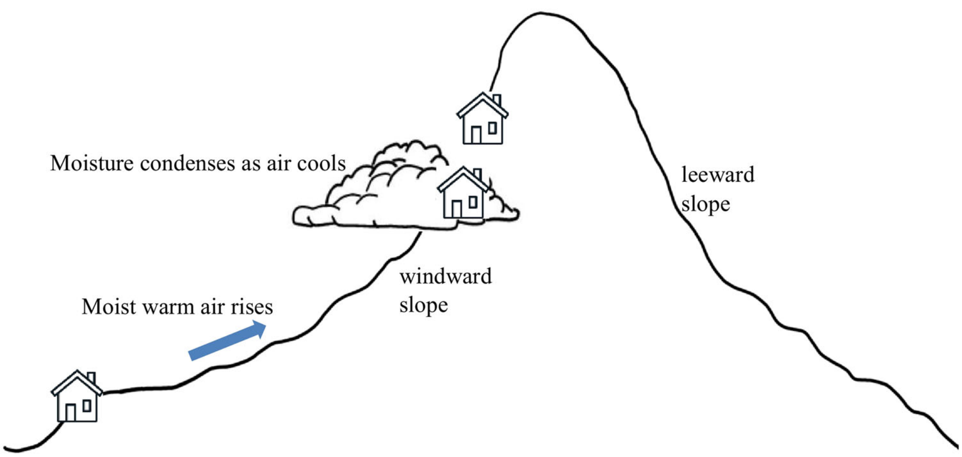



**Fig. 4 Cloud formation due to orographic lift.** Moist warm air is forced to rise when it runs into a topographic barrier. As the elevation increases and temperature goes down, moisture condenses into clouds.

and nimbostratus. According to US National Oceanic and Atmospheric Administration (NOAA) and Ahrens (2012, p. 105–106), the former, abbreviated as St, is a low greyish cloud layer with a fairly uniform base; at lowland, a stratus cloud often resembles a fog that does not touch the ground and fog is a surface-based form of stratus cloud. Normally, there is no precipitation falling from the stratus. The latter, abbreviated as Ns, is a dark gray, wet-looking cloud layer; it is often associated with more or less continuously falling rain or snow.

Therefore, frequent contact with low cloud suggests that it is not easy or not necessary for the Tibeto-Burman speakers to distinguish low cloud from fog. When low clouds occur in their highland environment, whose frequency is high (Wei et al., 2020), they have different experience with the clouds from people living near the sea level. Liu et al. (2018) point out that the major reason for low cloud formation in the Tibeto-Burman region, such as the southeast of the Qinghai-Tibet plateau, is due to orographic uplift. Orographic uplift is defined by NOAA as a phenomenon to occur when horizontally moving air is forced to rise before they go through a large obstacle, such as hills or mountains. The forced lifting due to the topographic barrier results in cooling, another important condition for cloud formation. If the air is humid and the cooling is sufficient, water vapor condenses into clouds. Due to orographic uplift, the low cloud may float on the mountaintop or just around the waist of the mountains. The residents who live there can treat the low cloud differently from the lowland people. While the lowland people see the low cloud above them, the mountain people often see the low cloud around them or beneath them (see Fig. 4).

Additionally, regarding the comparative non-Tibeto-Burman data, even though these languages are spoken in areas where the average relative humidity (74.29%) is higher than that of the Tibeto-Burman region, without the orographic uplift caused by the rising elevation, people's perception of low cloud can be completely different.

**Contact-induced fog-cloud similarity and divergence**. By looking at the proto-forms, some TB languages have maintained fog-cloud similarity (e.g., rGyalrongic languages) and divergence (e.g., Lolo-Burmese languages) for a long time. But some TB varieties display more recent changes through lexical borrowing. Due to the contact, they have gained or lost fog-cloud similarity or divergence. For example, while the other rGyalrongic languages keep using the PTB cloud morpheme *s-dim for both cloud and fog, some rGyalrongic varieties borrowed the fog word from Old Tibetan smug-pa and thus lost fog-cloud similarity. The fog word in rGyalrong (Aba Rongan Menggucun), rGyalrong (Maerkang R̪ibu), and r̪Gyalrong (Rangtang Puxicun) are smǝkpe, smǝkpe, and smǝkpa, while their cloud words are zdim, zdjǝm, and zdo, respectively (Nagano and Prins, 2013). Since fog and cloud are common weather phenomena, the borrowing occurs because of the prestige of the source language, rather than any need of naming new items. Within the Trans-Himalayan region, Tibetan culture is among the most influential ones, especially in the Tibeto-Burman area, hence the borrowing from Tibetan to rGyalrong. The Tibetan influence also reached non-Tibeto-Burman languages. For example, Tongren Bonan and Jishishan Bonan, two Mongolic varieties spoken in Qinghai and Gansu, China, both borrowed the words for fog and cloud from Amdo Tibetan, directly or indirectly. While Jishishan Bonan, with an elevation of 2485 m, spoken in Jishishan Bonan, Dongxiang and Salar Autonomous County, Linxia, Gansu, China, displays fog-cloud similarity, namely mokǝ 'cloud' and ɢɑdʐ̩ir mokǝ (ground cloud) 'fog' (Ding, 2022), Tongren Bonan, with an elevation of 1955 m, spoken in Tongren County, Huangnan Tibetan Autonomous Prefecture, Qinghai, China, differentiates ʂ̩ǝn 'cloud' from mukuɑ 'fog' (Bai, 2022). Due to the influence of Tibetan culture, different varieties of Bonan can either have fog-cloud similarity or fog-cloud divergence after borrowing from the prestigious language.

Other examples of borrowing concern another prestigious group of languages: the Sinitic languages. For example, while Bijiang Bai, a Northern Bai dialect with an elevation of 1808 m, spoken in Yunnan, China, colexifies fog and cloud, namely $\text{mɯ}^{21}\text{ko}^{42}$ 'fog, cloud' (CASS, 1991), Baishi Bai, another Northern Bai dialect with an elevation of 2278 m, spoken in Yunnan, lost fog-cloud similarity after borrowing the Chinese word $\text{y}^{35}$ from the local Southwest Mandarin: $\text{y}^{35}$ 'cloud' and $\text{mɯ}^{35}\text{kɔ}^{42}$ 'fog' (Yang, 2014). Furthermore, Lianghe Achang, a Burmish language in Dehong, Yunnan, China, with an elevation of 1301 m, gained fog-cloud similarity through language contact. It borrowed $\text{u}^{33}\text{lu}^{33}$ (fog:dew) from the local Mandarin to colexify fog and cloud; it is also fine to use $\text{u}^{33}$ without the dew morpheme for 'fog' in Lianghe Achang (Shi, 2009). In Chinese languages, it is common to use $\text{wu}^{51}\text{lu}^{51}$ (fog:dew) or its variants for 'fog', such as in Yantai Mandarin, Yudu Hakka, Danzhou Cantonese, Pingxiang Gan, and Ningbo Wu. Unlike Lianghe Achang, Luxi Achang, a close dialect of the former, with an elevation of 958 m, also borrowed $\text{u}^{55}\text{lu}^{35}$ from local Mandarin for 'fog', but does not replace its cloud word $\text{na}^{55}\text{mau}^{55}$ (sky:cloud) 'cloud' (Dai and Cui, 1985).

Our data also suggest that languages prefer differentiating once they have the linguistic and cultural impetus to do so. There are more contact-induced cases of fog-cloud divergence languages than of fog-cloud similarity languages in our samples. In other words, language contact chiefly promotes differentiation. This observation supports Regier et al.'s (2016) asymmetric pattern that there is a general preference for informative and precise communication.

**Application to CLICS data**. There are 183 cases of fog-cloud colexification in CLICS, including 33 TB languages. After we gained the necessary geospatial information (e.g., location and elevation) of the data in CLICS and removed the repetitive data points and all TB data, there are 131 varieties left, from 34 language families.

The average elevation of fog-cloud colexification data in CLICS is 983.3 m, lower than the TB data, but still much higher than the average elevation (526.4 m) of the fog-cloud divergence languages of our non-TB sample languages (see Table 5). This means that elevation remains to be a difference between languages of fog-cloud similarity and those of fog-cloud divergence. Our conclusion, namely fog-cloud similarity is more likely to occur at higher elevations, is supported by 46 languages/dialects in CLICS, or 35.1%, which are used at elevations ranging from 1000 m to 3000 m. The 46 languages are mainly from Austroasiatic, Camsá, Mpur, Kunza, Indo-European, Barbacoan, Nuclear Trans New Guinea, Austronesian, Timor-Alor-Pantar, and Daghestanian families. For example, the 34 Daghestanian languages stand out with an average elevation of 1758.1 m and a median elevation of 1713.5 m, spoken in the rugged mountainous Caucasus region.

However, while some Nuclear Trans New Guinea and Austronesian languages support our conclusion, which are spoken at high elevations, such as Kobon (2671 m) and Pazeh (2514 m), some are used at low elevations, such as Bima (15 m) and Apali (121 m). It seems to be a challenge to our conclusion that 51 languages/dialects of fog-cloud colexification are spoken below the elevation 500 m in CLICS (average 211 m), a range which is the least likely for fog-cloud similarity to occur, according to our TB and non-TB data. The table in Fig. 3 shows that only 4 languages in our sample displaying fog-cloud similarity are below elevation 500 m, all from the non-Tibeto-Burman samples. After we checked the distribution of the 51 languages/dialects from CLICS, 46 of them, or 90.2%, are located in East Nusa Tenggara (Indonesia), Timor-Leste (or East Timor), Papua New Guinea, and Amazon rainforest ecoregions (see Fig. 5).

These areas happen to feature tropical climates, characterized by year-long high temperatures, high humidity, and high precipitation (Beck et al., 2018; Galvin, 2016). Galvin (2016, p. 28) indicates that the cloudiest tropical zone stretches across the central Indian Ocean, Indonesia, and Malaysia to New Guinea. Therefore, rather than being a challenge to our conclusion, this observation of colexification below 500 m points to another probable environmental predictor for fog-cloud colexification: the tropical climate. This also explains the colexifying languages in the low elevations in Urban (2023), which are spoken in the Amazon rainforest ecoregions in South America (see Fig. 5).

Besides high humidity, the lowland tropical zone also has the condition to cool the water vapor, though not through orographic uplift as in the Tibeto-Burman region. Atkinson (2002) points out that stratus cloud is common along the tropical coasts where warm moist air is advected over cool coastal waters. After the stratus cloud is cooled, it may reach the water or ground surface. Moreover, advection fog can also be formed by warm moist air moving over a colder surface and cooling to its saturation point

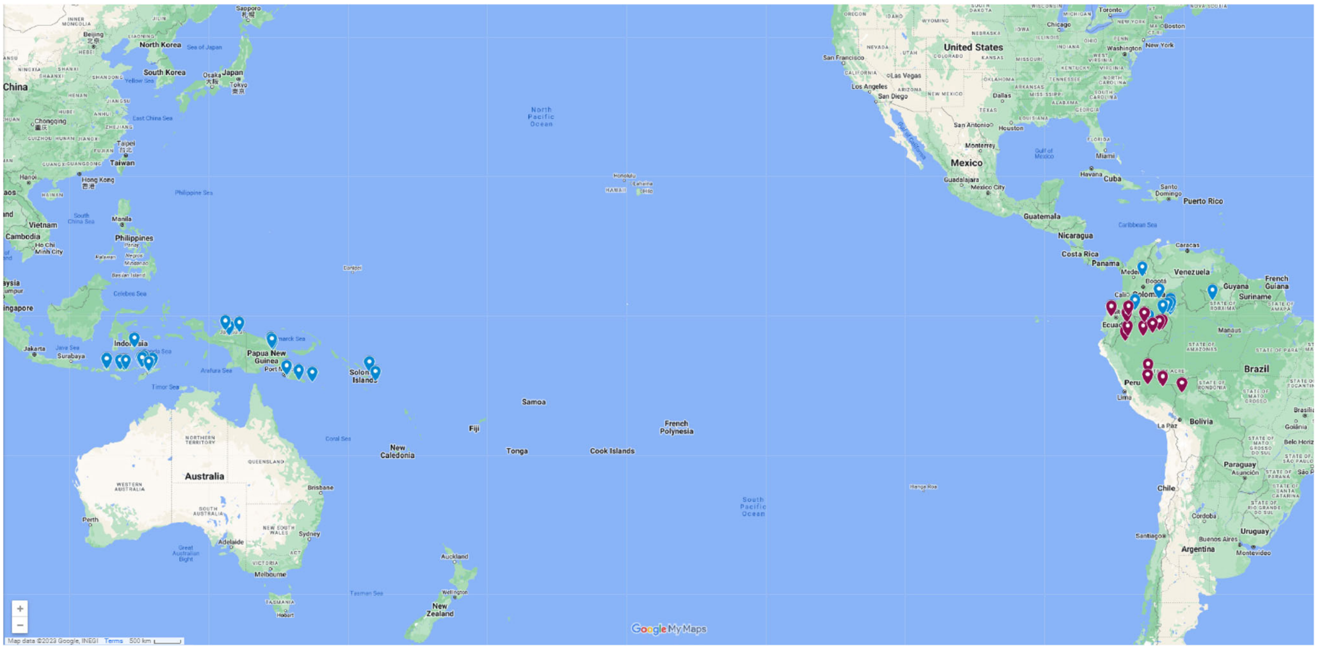



**Fig. 5 Fog-cloud colexification in tropical climates.** 46 languages/dialects from CLICS (in blue), and 14 languages in Urban (2023) (in purple), spoken lower than the elevation of 500 m, are located in the tropical regions, namely East Nusa Tenggara (Indonesia), Timor-Leste (or East Timor), Papua New Guinea, and the Amazon rainforest ecoregions.

(Ahrens, 2012, p. 98). This kind of environment provides the cognitive conditions for people to mix low cloud with fog. This may explain the fog-cloud colexification in languages along the coasts of East Nusa Tenggara and Timor-Leste.

Papua New Guinea and the Amazon basin also belong to the tropical zone. But they have a tropical rainforest climate, different from the tropical savanna climate of East Nusa Tenggara and Timor-Leste (Beck et al., 2018), resulting in a different mechanism for cloud/fog formation. The trees and other plants in the rainforest transpire vast amounts of water vapor from their leaves and release tiny particles serving as cloud condensation nuclei, around which water droplets condense to form clouds and eventually rain (Pöhlker et al., 2012; Fenning, 2014). According to Obregon et al. (2014), lowland rainforests also feature frequent occurrence of ground-touching clouds, which are in contact with the forest canopy and are perceived as fog at the surface. Therefore, due to the frequent formation of fog/low stratus cloud, this type of rainforest is called "tropical lowland cloud forest" (Gradstein et al., 2010; Obregon et al., 2011; Gehrig-Downie et al., 2012). Interestingly, since fog and cloud are very hard to distinguish in tropical lowland rainforests, Obregon et al. (2014, p. 322) propose the use of the term "lowland fog forest" as a synonym for "lowland cloud forest".

In sum, cases of lowland fog-cloud similarity, specifically fog-cloud colexification, in the database of CLICS and Urban (2023), do not contradict our conclusion by the Tibeto-Burman languages. On the one hand, many colexification languages in CLICS support our conclusion. On the other hand, those which do not corroborate are actually pointing to another predictor for fog-cloud similarity, i.e., the tropical climate. It is worth future investigation with expanded sample languages in the tropical zone.

## Conclusions

The goal of the present study is to investigate the influence of natural environment upon linguistic expressions, specifically the influence of elevation upon the lexical use of fog and cloud in Tibeto-Burman languages. After studying 234 Tibeto-Burman languages/dialects and comparing them with 213 non-Tibeto-Burman languages in the Trans-Himalayan region, it is found that more than half of the Tibeto-Burman languages display fog-cloud similarity, and it is more likely to happen at higher elevations, particularly between the range of 1000 to 3000 m. The high proportion (i.e., 52.99%) of fog-cloud similarity in Tibeto-Burman languages, compared with that of the non-Tibeto-Burman languages (i.e., 10.80%), shows that languages are adaptive to ecological conditions.

There are three lexical relations for fog-cloud similarity in Tibeto-Burman languages. While some Tibeto-Burman languages colexify fog and cloud, some consider fog a hyponym of cloud, using the cloud morpheme as the head with other modificatory morphemes. In some other Tibeto-Burman languages, although fog is expressed with a different morpheme or related to a different concept (e.g., ash, dew, smoke), cloud must be a formative of the fog expression, though not as the head; in other words, cloud is part of the fog. The other half of the Tibeto-Burman languages use semantically disconnected words to describe fog and cloud.

After reviewing the meteorological features, we found that the Tibeto-Burman region has the ideal conditions for the formation of low cloud, mainly the stratus and nimbostratus cloud. Firstly, it is very humid. Secondly, its topography can cool the moist air. When the horizontally moving moist air runs into the topographic barrier, the high elevation forces it to rise and cool, and the moist air eventually condenses into clouds, a process called orographic uplift. Since Tibeto-Burman speakers live in high elevations, low cloud, the dominant cloud of the region, may surround them or beneath their view. Therefore, they may find it difficult or not necessary to distinguish fog from low cloud.

Moreover, our findings support Regier et al.'s (2016) theory of efficient communication. The fog-cloud similarity languages, including both strict and loose colexification, are more constrained than the fog-cloud divergence languages with regard to the non-linguistic variable, namely elevation in the present study. It suggests that languages displaying fog-cloud similarity are adaptive to higher elevations with lower communicative need to distinguish between the two concepts by using completely different and unrelated linguistic forms. On the contrary, fog-cloud divergence languages have stronger need, resulting from the physical environment, to communicate by using completely different concepts and thus different linguistic forms.

Furthermore, we have identified other factors than the physical environment, playing their roles in the lexical use of "fog" and "cloud" among the Tibeto-Burman languages, namely the lineage-specific preference, and the effect of language contact. At the lower nodes of the family tree, some closely related varieties can, not necessarily though, display the lineage-specific effect, such as the Tibetan. But the lineage-specific effect is not found at higher nodes of the family tree. Contact-induced cases of fog-cloud similarity and divergence are also found. After borrowing from prestigious languages (e.g., Tibetan and Chinese), close dialects or varieties can behave differently regarding their lexical use of fog and cloud. Meanwhile, language contact promotes differentiation since there are more contact-induced cases of fog-cloud divergence than of fog-cloud similarity in our samples. The result is confirmative of Regier et al.'s (2016) asymmetric pattern, which suggests that there is a general preference for informative and precise communication.

Therefore, the causal link between higher elevation and fog-cloud similarity should not be treated as deterministic, but probabilistic. Parallel to Regier et al.'s (2016) findings based on ice and snow, not all languages at high elevations will necessarily collapse the fog and cloud distinction. A probabilistic stance indicates that there is less communicative need to preserve the distinction between fog and cloud at higher elevations and there is higher communicative need to distinguish them at lower elevations.

Finally, our conclusion, namely fog-cloud similarity is more likely to occur between the elevation 1000 and 3000 m, is supported by 46 languages/dialects, or 35.1%, in CLICS. Instead of being a challenge to our conclusion, the CLICS data and Urban's (2023) samples of lowland languages below elevation 500 m point to another predictor for fog-cloud similarity, i.e., the tropical climate, which is a direction for future investigation.

## Data availability

The datasets generated during and/or analyzed during the current study are available in the Dataverse repository: https://doi.org/10.7910/DVN/S6PTEJ.

Received: 30 July 2022; Accepted: 20 June 2023;
Published online: 03 July 2023

## Notes

1 CLICS uses the term "Sino-Tibetan". But since there are no Sinitic languages in this database, the "Sino-Tibetan" languages in CLICS are all Tibeto-Burman.
2 Similar reconstructions to PTB *bʷar/*pʷar 'fire' in Matisoff (2003) are *bwár / *pwár in Benedict (1972) and *bar in Coblin (1986). The reflexes of this etymon are mainly verbs, such as par 'to burn' in Apatani (Macro-Tani) (Sun, 1993), and bar 'to burn (intransitive)' and par 'to burn (transitive)' in Kanauri (Bodic) (Benedict, 1972). But many of its reflexes in Brahmaputran languages are nouns, such as wan[31] 'fire' in Jingpho (Liu, 1984) and wal 'fire' in Garo (Burling, 2003).
3 VanBik (2009) and Mortensen (2012) reconstructed 'cloud, fog' of Proto-Kuki-Chin as *may and of Proto-Tangkhulic as *moj, respectively, consistently featuring the nucleus as complex vowels.
4 Matisoff's (2003) reconstruction *r-məw is similar to Benedict's (1972) *(r-)muw and Weidert's (1987) *(r-)məw / *(r-)muw.
5 For example, many Loloish (or Ngwi) cloud words are descendants of *s-dim or Proto-Loloish *C-dim[1], reconstructed by Bradley (1979).
6 Accessed at https://stedt.berkeley.edu/.
7 Accessed at https://htq.minpaku.ac.jp/databases/rGyalrong/lang/index.php?langindex=eng.
8 The Chinese name of the database is 中国语言资源保护工程采录展示平台. Accessed at https://zhongguoyuyan.cn/index.html?lang=cn.
9 Accessed at https://abvd.eva.mpg.de/austronesian/.
10 Guiyang, the capital city of Guizhou Province, southwest China, has an average annual relative humidity of 77 % (Chen et al., 2021) and was the most humid city in China in 2020, according to https://www.statista.com/statistics/282491/china-annual-average-humidity-in-major-cities/.
11 According to Yang (2010), mi[55] should be a reflex of PTB *muːŋ 'foggy, fog'.
12 According to Ring (2015), when both formatives are nouns, noun compounds in Pnar are functionally genitival expressions, such as ka=balaŋ lathadlabɔt 'Lathadlabot church' or 'church of Lathadlabot', where 'church' is the head.
13 There is another possible, but less likely, interpretation of the semantic formation of 'fog' in Naic languages, namely literally "cloud smoke". Although it is possible for the colexification to occur between 'fog' and 'smoke' (see section "Fog is 'cloud smoke'"), typologically, the connection between 'firewood' and 'smoke' is not attested. Due to a lack of intermediate connection between the two meanings, therefore, we are more confident to propose the semantic development from 'firewood' to 'ash, dust'.
14 In Pumi, nasal vowels are very widespread, some of which originated in nasal codas (Michaud, Jacques and Rankin, 2012).

## References

Ahrens CD (2012) Essentials of meteorology: an invitation to the atmosphere, 6th edn. Brooks/Cole, Belmont
Atkinson GD (2002) Forecasters' guide to tropical meteorology. University Press of the Pacific, Honolulu HI
Axelsen JB, Manrubia S (2014) River density and landscape roughness are universal determinants of linguistic diversity. Proc Biol Sci 281(1784):1–8. https://doi.org/10.1098/rspb.2013.3029
Baddeley R, Attewell D (2009) The relationship between language and the environment: information theory shows why we have only three lightness terms. Psychol Sci 20:1100–1107
Bai A (2022) Qinghai Tongren Bao'anyu (Tongren dialect of Bonan in Qinghai). https://zhongguoyuyan.cn/point/65779. Accessed 23 Dec 2022
Bates E, MacWhinney B (1982) Functionalist approaches to grammar. In: Wanner E, Gleitman L (eds) Language acquisition: the state of the art. Cambridge University Press, New York, pp. 173–218
Beck H, Zimmermann N, McVicar T et al. (2018) Present and future Köppen-Geiger climate classification maps at 1-km resolution. Sci Data 5:180214. https://doi.org/10.1038/sdata.2018.214
Benedict PK (1972) Sino-Tibetan: a conspectus. Cambridge University Press, New York
Benedict PK (1976) Rhyming dictionary of written Burmese. Linguist Tibeto-Burman Area 3(1):1–93
Boas F (1911) Introduction. In: Handbook of American Indian languages, Vol 1. Government Print Office (Smithsonian Institution, Bureau of American Ethnology, Bulletin 40), pp, 1–83
Du Bois JA (1985) Competing motivations. In: Haiman J (ed.) Iconicity in syntax. John Benjamins Publishing Company, Amsterdam, pp. 343–366
Bradley D (1979) Proto-Loloish. Curzon Press, London and Malmö
Burenhult N, Levinson SC (2008) Language and landscape: a cross-linguistic perspective. Lang Sci 30(2–3):135–150. https://doi.org/10.1016/j.langsci.2006.12.028
Burling R (1983) The Sal languages. Linguist Tibeto-Burman Area 7(2):1–32
Burling R (2003) The language of the Modhupur Mandi (Garo) Vol. III: glossary. University of Michigan, Ann Arbor
De Busser R (2015) The influence of social, cultural, and natural factors on language structure: an overview. In: De Busser R, LaPolla RJ (eds) Cognitive linguistic studies in cultural contexts. John Benjamins Publishing Company, Amsterdam, pp. 1–28
Carling G (2017) Diachronic atlas of comparative linguistics online. https://diacl.ht.lu.se/. Accessed 12 May 2022
CASS (or Chinese Academy of Social Sciences) (1991) Zangmianyu yuyin he cihui (Tibeto-Burman phonology and lexicon). China Social Sciences Press, Beijing
Chang H (1986) Lahuyu jianzhi (A concise grammar of Lahu). The Ethnic Publishing House, Beijing
Chen X, Wang Z, Bao Y (2021) Cool island effects of urban remnant natural mountains for cooling communities: a case study of Guiyang, China. Sustain Cities Soc 71:102983. https://doi.org/10.1016/j.scs.2021.102983
Chou F (1972) Archaic Chinese and Sino-Tibetan. J Chin Stud Chin Univ Hong Kong 5(1):159–237
Coblin WS (1986) A sinologist's handlist of Sino-Tibetan lexical comparisons. Steyler Verlag, Nettetal
Coupé C, Maddieson I (2016) Quelle adaptation acoustique pour les langues du monde? In: Actes du 13ème congrès Français d'acoustique, Université du Maine, Le Mans (France), 11–15 April 2016

Croft W (2003) Typology and universals. Cambridge University Press, Cambridge, UK
Dai QX, Cui ZC (1985) Achangyu jianzhi (A sketch of Achang). The Ethnic Publishing House, Beijing
Day JA (1998) Fog and mist. In: Herschy RW (ed) Encyclopedia of Hydrology and Lakes. Springer, Dordrecht
Ding SZ (2014) A grammar of Prinmi: based on the central dialect of northwest Yunnan, China. Brill, Leiden
Ding SQ (2022) Jishishan Bao'anyu (Jishishan dialect of Bonan). https://zhongguoyuyan.cn/point/65603. Accessed 23 Dec 2022
Domrös M, Peng G (1988) The climate of China. Springer, Berlin and Heidelberg
Dong S, Huang C-R, Ren H (2020) Towards a new typology of meteorological events: a study based on synchronic and diachronic data. Lingua 247:102894
Dong S, Yang Y, Ren H, Huang C-R (2021) Directionality of atmospheric water in Chinese: a lexical semantic study based on linguistic ontology. SAGE Open 11:1
Van Driem G (1993) A grammar of Dumi. Mouton de Gruyter, Berlin
Van Driem G (2007) The diversity of the Tibeto-Burman language family and the linguistic ancestry of Chinese. Bull Chin Linguist 1(2):211–270
Dryer MS (2008) Word order in Tibeto-Burman languages. Linguist Tibeto-Burman Area 31(1):1–83
Dunn M, Greenhill S, Levinson S et al. (2011) Evolved structure of language shows lineage-specific trends in word-order universals. Nature 473:79–82. https://doi.org/10.1038/nature09923
Ember CR, Ember M (2010) Climate, econiche, and sexuality: influences on sonority in language. Am Anthropol 109:180–185
Evans JP (1999) Introduction to Qiang phonology and lexicon: synchrony and diachrony. Dissertation. University of California, Berkeley
Everett C (2017) Languages in drier climates use fewer vowels. Front Psychol 8:1285. https://doi.org/10.3389/fpsyg.2017.01285
Everett C, Blasi DE, Roberts SG (2015) Climate, vocal folds, and tonal languages: connecting the physiological and geographic dots. Proc Natl Acad Sci USA 112(5):1322–1327. https://doi.org/10.1073/pnas.1417413112
Fenning T (2014) Challenges and opportunities for the world's forests in the 21st century. Springer, Dordrech
Finley S (2018) Cognitive and linguistic biases in morphology learning. Wiley Interdiscip Rev 9(5):e1467. https://doi.org/10.1002/wcs.1467
Fought JG, Munroe RL, Fought CR, Good EM (2004) Sonority and climate in a world sample of languages. Cross-Cult Res 38:27–51. https://doi.org/10.1177/1069397103259439
François A (2008) Semantic maps and the typology of colexification: intertwining polysemous networks across languages. In: Vanhove M (ed) From polysemy to semantic change: towards a typology of lexical semantic associations. John Benjamins Publishing, Amsterdam, pp. 163–215
Gabelentz G (1901) Die sprachwissenschaft: ihre aufgaben, methoden und bisherigen ergebnisse. Tauchnitz, Leipzig
Galvin JFP (2016) An introduction to the meteorology and climate of the tropics. John Wiley & Sons, Chichester
Gao Y (2022) Sichuan Kangding Muyayu xibu fangyan. Western dialect of Muya in Kangding, Sichuan, https://zhongguoyuyan.cn/point/60773 Accessed 23 Dec 2022
Gehrig-Downie C, Marquardt J, Obregón A, Bendix J, Gradstein SR (2012) Diversity and vertical distribution of filmy ferns as a tool for identifying the novel forest type "tropical lowland cloud forest". Ecotropica 18(1):35–44
Genga WM (2019) Sichuan Daofu Ergongyu (A grammar of Ergong in Sichuan). Commercial Press, Beijing
Glover WW (1972) A vocabulary of the Gurung language. Summer Institute of Linguistics and Institute of Nepal Studies, Tribhuvan University, Kirtipur
Gong QH (2007) Zhabayu yanjiu (A grammar of Zhaba). The Ethnic Publishing House, Beijing
Gorenflo LJ, Romaine S, Mittermeier RA, Walker-Painemilla K (2012) Co-occurrence of linguistic and biological diversity in biodiversity hotspots and high biodiversity wilderness areas. Proc Natl Acad Sci USA 109(21):8032–8037
Gradstein SR, Obregon A, Gehrig C, Bendix J (2010) Tropical lowland cloud forest: A neglected forest type. In: Bruijnzeel LA, Scatena FN, Hamilton LS (eds) Tropical montane cloud forests: science for conservation and management. Cambridge University Press, Cambridge, pp. 130–133
Greenhill S, Gray R (2015) Bantu Basic Vocabulary Database. https://clics.clld.org/languages/bantubvd-2. Accessed 5 Jan 2023
Haiman J (2010) Competing motivations. In: Song JJ (ed.) The Oxford handbook of linguistic typology. Oxford University Press, Oxford, pp. 148–165
Hale A (1973) Clause, sentence, and discourse patterns in selected languages of Nepal IV: word lists. Summer Institute of Linguistics and Tribhuvan University Press, Kathmandu
Hammarström H, Forkel R, Haspelmath M, Bank S (2022) Glottolog 4.6. Max Planck Institute for Evolutionary Anthropology, Leipzig
Haspelmath M, Tadmor U (2009) World loanword database. Max Planck Institute for Evolutionary Anthropology, Leipzig
He JR, Jiang ZY (1985) Naxiyu jianzhi (A concise grammar of Naxi). The Ethnic Publishing House, Beijing
Honkasalo S (2019) A grammar of eastern Geshiza: a culturally anchored description. Dissertation, University of Helsinki
Hoshi M (1984) A Prakaa vocabulary: a dialect of the Manang language. Anthropological and linguistic studies of the Gandaki Area in Nepal II. ILCAA, Tokyo
Huang BF (1992) Zangmianyuzu yuyan cihui (A Tibeto-Burman lexicon). Central Institute of Minorities, Beijing
Huang C-R, Dong S, Yang Y, Ren H (2021) From language to meteorology: kinesis in weather events and weather verbs across Sinitic languages. Humanit Soc Sci Commun 8:4
Huang Y (2020) Zhabayu de mingwuhua he guanxihua (Nominalization and relativization in nDrapa). Minzu Yuwen (Minority languages of China) 4:29–42
Jacques G (2015) On the cluster *sr– in Sino-Tibetan. J Chin Linguist 43(1):215–223
Jacques G (2021) A grammar of Japhug. Language Science Press, Berlin
Jacques G, Michaud A (2011) Approaching the historical phonology of three highly eroded Sino-Tibetan languages: Naxi, Na and Laze. Diachronica 28:468–498
Jacques G (forthcoming) An overview of morphology in Sino-Tibetan/Trans-Himalayan. In: Arkadiev P, Rainer F (eds) The Oxford handbook of historical morphology. Oxford University Press, Oxford
Jacquesson F (2008) A Kokborok grammar (Agartala dialect). Kókborok Tei Hukumu Mission, Agartala
Jaswal AK, Kore PA, Singh V (2017) Variability and trends in low cloud cover over India during 1961-2010. MAUSAM 68(2):235–252
Jiang Y (2015) Dayang pumiyu cankao yufa (A grammar of Dayang Prinmi). China Social Sciences Press, Beijing
Kaiping GA, Edwards O, Klamer M (2019) LexiRumah 3.0.0. https://lexirumah.model-ling.eu/. Accessed 12 May 2022
Kattel DB, Yao T, Yang K, Tian L, Yang G, Joswiak D (2013) Temperature lapse rate in complex mountain terrain on the southern slope of the central Himalayas. Theor Appl Climatol 113:671–682
Key MR, Comrie B (2015) The intercontinental dictionary series. Max Planck Institute for Evolutionary Anthropology, Leipzig
Konnerth L (2016) The Proto-Tibeto-Burman *gV-nominalizing prefix. Linguistics of the Tibeto-Burman Area 39(1):3–32
Kurosaki Y, Kimura F (2002) Relationship between topography and daytime cloud activity around Tibetan Plateau. J Meteoroll Soc Jpn 80(6):1339–1355
Lai YF (2017) Grammaire du khroskyabs de Wobzi. Dissertation, Université Sorbonne Paris Cité
LaPolla RJ (1987) Dulong and Proto-Tibeto-Burman. Linguist Tibeto-Burman Area 10(1):1–43
LaPolla RJ, Huang CL (2003) A grammar of Qiang: with annotated texts and glossary. De Gruyter, Mouton, Berlin, New York
Levinson SC (2003) Space in language and cognition: explorations in cognitive diversity. Cambridge University Press, Cambridge
Levinson SC, Wilkins DP (2006) Grammars of space: explorations in cognitive diversity. Cambridge University Press, Cambridge
Li FW (1997) Xiahan zidian (Tangut-Chinese Dictionary). China Social Sciences Press, Beijng
Li C (2022a) Yunnan Ninglang Lisuyu Ninglang fangyan cuiyuhua (Cuiyu dialect of Ninglang Lisu in Yunnan). https://zhongguoyuyan.cn/point/60M40. Accessed 23 Dec 2022
Li C (2022b) Sichuan Dechang Lisuyu Nujiang fangyan Dechanghua (Dechang Dialect of Lisu). https://zhongguoyuyan.cn/point/60826. Accessed 23 Dec 2022
Li C (2022c) Yunnan Yongsheng Lisuyu Nujiang fangyan Liudetuyu (Yongsheng dialect of Lisu). https://zhongguoyuyan.cn/point/60717. Accessed 23 Dec 2022
Liu CH (1995) Liuzhou fanyan cidian (A dictionary of Liuzhou Mandarin). Jiangsu Education Press, Nanjing
Liu GK (1998) Mawo Qiangyu yanjiu (A grammar of Mawo Qiang). Sichuan Ethnic Publishing House, Chengdu
Liu L (1984) Jingpozu yuyan jianzhi (A concise grammar of the Jingpho language). The Ethnic Publishing House, Beijing
Liu YM, Yan YF, Lü JH, Liu XL (2018) Review of current investigations of cloud, radiation and rainfall over the Tibetan Plateau with the CloudSat/CALIPSO dataset. Chin J Atmos Sci (in Chinese) 42(4):847–858. https://doi.org/10.3878/j.issn.1006-9895.1805.17281
Liu ZF (2022) (2022) Sichuan Yanbian Naxiyu dongbu fanyan Mosuohua. Mosuo speech in Yanbian, Sichuan, https://zhongguoyuyan.cn/point/60M18 Accessed 23 Dec 2022
Lu SZ (1983) Pumiyu jianzhi (A concise grammar of Prinmi). The Ethnic Publishing House, Beijing
Lu SZ (2001) Pumiyu fangyan yanjiu (A study of Prinmi dialects). The Ethnic Publishing House, Beijing

Maddieson I (2018) Language adapts to environment: sonority and temperature. Front Commun 3:28
Maddieson I, Coupé C (2015) Human spoken language diversity and the acoustic adaptation hypothesis. J Acoust Soc Am 138(3):1838–1838. https://doi.org/10.1121/1.4933848
Maddieson I (2012) On the origin and distribution of complexity in phonological structure. In: Proceedings of the joint conference JEP-TALN-RECITAL 2012. p7. Available online at: http://aclweb.org/anthology/F/F12/F12-4004.pdf
Marrison GE (1967) The classification of the Naga languages of north-east India (Volume 2). Dissertation, University of London
Martin L (1986) Eskimo words for snow: a case study in the genesis and decay of an anthropological example. Am Anthropol 88:418–423
Matisoff JA (1990) On megalocomparison. Language 66(1):106–120
Matisoff JA (1991) The mother of all morphemes: augmentatives and diminutives in areal and universal perspective. In: Ratliff M, Schiller E (eds.). Papers from the First Annual Meeting of the Southeast Asian Linguistic Society. Arizona State University, Tempe
Matisoff JA (2003) Handbook of Proto-Tibeto-Burman: system and philosophy of Sino-Tibetan reconstruction. University of California Press, Berkeley
Matisoff JA (2006) English-Lahu lexicon. University of California Press, Berkeley
Michaud A, Jacques G, Rankin RL (2012) Historical transfer of nasality between consonantal onset and vowel: from C to V or from V to C? Diachronica 29(2):201–230
Michaud A (2018) Na (Mosuo)-English-Chinese dictionary. https://halshs.archives-ouvertes.fr/halshs-01204638v3/document. Accessed 22 Sept 2020
Mortensen DR (2012) Database of Tangkhulic languages. http://stedt.berkeley.edu/search/. Accessed 9 Dec 2022
Mu YZ, Sun HK (2012) Lisuyu fangyan yanjiu (A study of Lisu dialects). The Ethnic Publishing House, Beijing
Muellner-Riehl AN (2019) Mountains as evolutionary arenas: patterns, emerging approaches, paradigm shifts, and their implications for plant phylogeographic research in the Tibeto-Himalayan region. Front Plant Sci 10:195. https://doi.org/10.3389/fpls.2019.00195
Munroe RL, Silander M (1999) Climate and the consonant-vowel (CV) syllable: replication within language families. Cross-Cult Res 33:43–62. https://doi.org/10.1177/106939719903300104
Munroe RL, Munroe RH, Winters S (1996) Cross-cultural correlates of the consonant-vowel syllable. Cross-Cult Res 30:60–83. https://doi.org/10.1177/106939719603000103
Munroe RL, Fought JG, Macaulay RKS (2009) Warm climates and sonority classes: not simply more vowels and fewer consonants. Cross-Cult Res 43:123–133. https://doi.org/10.1177/1069397109331485
Nagano Y (1984) A Manang glossary. Anthropological and linguistic studies of the Gandaki Area in Nepal II. ILCAA, Tokyo
Nagano Y (2017) Cogtse rGyarong. In: Thurgood G, LaPolla RJ (eds) The Sino-Tibetan languages. Routledge, London
Nagano Y, Prins M (2013) rGyalrongic languages database. https://htq.minpaku.ac.jp/databases/rGyalrong/. Accessed 20 May 2022
Nagaraja KS, Sidwell P, Greenhill S (2013) A lexicostatistical study of the Khasian languages. Mon-Khmer Studies 42:1–11
Nichols J (1992) Linguistic diversity in space and time. University of Chicago Press, Chicago
Nogués-Bravo D, Araújo MB, Romdal T, Rahbek C (2008) Scale effects and human impact on the elevational species richness gradients. Nature 453(7192):216–219. https://doi.org/10.1038/nature06812
O'Meara C, Pérez Báez G (2011) Spatial frames of reference in Mesoamerican languages. Lang Sci 33(6):837–852. https://doi.org/10.1016/j.langsci.2011.06.013
Obregon A, Gehrig-Downie C, Gradstein SR, Bendix J (2014) The potential distribution of tropical lowland cloud forest as revealed by a novel MODIS-based fog/low stratus night-time detection scheme. Remote Sens Environ 155:312–324. https://doi.org/10.1016/j.rse.2014.09.005
Obregon A, Gehrig-Downie C, Gradstein SR, Rollenbeck R, Bendix J (2011) Canopy level fog occurrence in a tropical lowland forest of French Guiana as a prerequisite for high epiphyte diversity. Agric Forest Meteorol 151(3):290–300. https://doi.org/10.1016/j.agrformet.2010.11.003
Ouyang JY (1985) Luobazu yuyan jianzhi (A concise grammar of Bokar). The Ethnic Publishing House, Beijing
Palmer B (2015) Topography in language: absolute frame of reference and the topographic correspondence hypothesis. In: De Busser R, LaPolla RJ (eds) Cognitive linguistic studies in cultural contexts. John Benjamins Publishing Company, Amsterdam, pp. 177–226
Pöhlker C, Wiedemann KT, Sinha B et al. (2012) Biogenic potassium salt particles as seeds for secondary organic aerosol in the Amazon. Science 337(6098):1075–8. https://doi.org/10.1126/science.1223264
Popovich HA, Popovich FB (2005) Maxakalí-English dictionary. Summer Institute of Lingusitics, Cuiabá
Prins M (2016) A grammar of rGyalrong, Jiaomuzu (kyom-kyo) dialects: a web of relations. Brill, Leiden
Pullum G (1991) The great Eskimo vocabulary hoax and other irreverent essays on the study of language. University of Chicago Press, Chicago and London
Qumu, TX (2022) Sichuan Mianning Yiyu beibu fanyan Shuitianhua (Shuitian dialect of Liangshan Yi). https://zhongguoyuyan.cn/point/60832. Accessed 23 Dec 2022
Regier T, Kemp C, Kay P (2015) Word meanings across languages support efficient communication. In: MacWhinney B, O'Grady W (eds) The handbook of language emergence. Wiley-Blackwell, Hoboken, pp. 237–263
Regier T, Carstensen A, Kemp C (2016) Languages support efficient communication about the environment: words for snow revisited. PLoS ONE 11:e0151138
Ring H (2015) A grammar of Pnar. Dissertation, Nanyang Technological University
Rosch E (1999) Principles of categorization. In: Margolis E, Laurence S (eds.) Concepts: core readings. MIT Press, Cambridge, MA, pp. 189–206
Rzymski C, Tresoldi T, Greenhill SJ et al. (2020) The database of cross-linguistic colexifications, reproducible analysis of cross-linguistic polysemies. Sci Data 7(1):13. https://doi.org/10.1038/s41597-019-0341-x
Sagart L, Jacques G, Lai YF, Ryder R, Thouzeau V, Greenhill SJ, List JM (2019) Dated language phylogenies shed light on the history of Sino-Tibetan. Proc Natl Acad Sci USA 116:10317–10322
Sapir E (1912) Language and environment. Am Anthropol 14:226–242
Shi J (2009) Lianghe Achangyu cankao yufa (A grammar of Lianghe Achang). China Social Sciences Press, Beijing
Shi S (2018) Ethnic flows in the Tibetan-Yi corridor throughout history. Int J Anthropold Ethno 2:1–22
So-Hartmann H (1988) Notes on the Southern Chin languages. Linguist Tibeto-Burman Area 11(2):98–119
Sun TS (2014) Typology of generic-person marking in Tshobdun Rgyalrong. In: Simmons RV, Van Auken NA (eds) Studies in Chinese and Sino-Tibetan linguistics: dialect, phonology, transcription and text. Institute of Linguistics, Academia Sinica, Taipei
Sun TS (1993) Tani synonym sets. http://stedt.berkeley.edu/search/. Accessed 13 Jul 2022
Tournadre N et al. (2009) Sherpa-English and English-Sherpa dictionary. Vajra Publications, Kathmandu
Urban M (2012) Analyzability and semantic associations in referring expressions: a study in comparative lexicology. Dissertation, Max Planck Institute for Evolutionary Anthropology, Leipzig
Urban M (2023) Foggy connections, cloudy frontiers: on the (non-)adaptation of lexical structures. Front Psychol 14:1115832
VanBik K (2009) Proto-Kuki-Chin: a reconstructed ancestor of the Kuki-Chin languages. STEDT, Berkeley
Walcek CJ (1994) Cloud cover and its relationship to relative humidity during a springtime midlatitude cyclone. Mon Weather Rev 122(6):1021–1035
Wang P (1994) Guiyang fangyan cidian (Guiyang Mandarin lexicon). Jiangsu Education Press, Jiangsu
Wei J, Duan KQ, Xin R (2020) Cloud occurrence probability and its radiative forcing characteristics in Qinghai-Tibet Plateau. J Glaciol Geocryol 42(2):368–377
Weidert A (1987) Tibeto-Burman tonology: a comparative account. John Benjamins Publishing Company, Amsterdam and Philadelphia
Wen J (2022) Yunnan Mangshi Zhongshan Langsuyu (Mangshi Zhongshan Dialect of Maru). https://zhongguoyuyan.cn/point/60855. Accessed 23 Dec 2022
Witkowski SR, Brown CH (1985) Climate, clothing, and body-part nomenclature. Ethnology 24:197–214
Xu Y, Duong K, Malt BC, Jiang S, Srinivasan M (2020) Conceptual relations predict colexification across languages. Cognition 201:104280–104280. https://doi.org/10.1016/j.cognition.2020.104280
Yan RX, Wong SW (1988) Pumizu jianshi (A sketch of the Prinmi nationality). Yunnan People Publisher, Kunming
Yang C (2010) Lalo regional varieties: phylogeny, dialectometry, and sociolinguistics. Dissertation, La Trobe University
Yang XX (2014) Baiyu Baishihua cankao yufa (A reference grammar of Baishi Bai language). Dissertation, Xiamen University
Zhan BH, Chen XJ, Li R (1997) Dongguan fangyan cidian (A lexicon of Dongguan Yue). Jiangsu Education Publishing House, Nanjing
Zhang JC (1986) Cangluo Menbayu jianzhi (Brief description of the Cangluo Menba language). The Ethnic Publishing House, Beijing
Zhang MH, Yan S, Pan WY, Jin L (2019) Phylogenetic evidence for Sino-Tibetan origin in northern China in the Late Neolithic. Nature 569:112–115. https://doi.org/10.1038/s41586-019-1153-z
Zhang SY (2020) Le rgyalrong situ de Brag-bar et sa contribution à la typologie de l'expression des relations spatiales: l'orientation et le mouvement associé. Dissertation, Institut National des Langues et Civilisations Orientales

Zhao QL (2022) Yunnan Lijiang Naxiyu xibu fanyan Dayanhua (Dayan Naxi in Lijiang, Yunnan). https://zhongguoyuyan.cn/point/60852. Accessed 23 Dec 2022

Zheng WX (2016) A grammar of Longxi Qiang. Dissertation, National University of Singapore

Zhou MC (2004) Maqu zangyu yanjiu (Grammar of Maqu Tibetan). The Ethnic Publishing House, Beijing

Zhou CF (2019) Re'ela Qiangyu cankao yufa (A grammar of Re'ela Qiang). Dissertation, Shanghai Normal University

## Acknowledgements

For valuable feedback on the linguistic data of this work, we thank Dr. Cathryn Yang, Dr. Sizhi Ding, Dr. Yang Huang, and Dr. Yunfan Lai.

## Author contributions

HDD: conception and design of study; data collection, analysis and interpretation; drafting the manuscript, revising the manuscript critically for important intellectual content; approval of the version of the manuscript to be submitted; agreement to be accountable for all aspects of the work. SCD: conception and design of study; data collection, analysis and interpretation; drafting the manuscript, revising the manuscript critically for important intellectual content; approval of the version of the manuscript to be submitted; agreement to be accountable for all aspects of the work.

## Competing interests

The authors declare no competing interests.

## Ethical approval

This article does not contain any studies with human participants performed by any of the authors.

## Informed consent

This article does not contain any studies with human participants performed by any of the authors.

## Additional information

**Correspondence** and requests for materials should be addressed to Hongdi Ding or Sicong Dong.