



---

## **GEOGRAPHICAL VARIATION AMONG THE THARU OF LUMBINI PROVINCE**

**Tara Mani Rai, PhD**

Assistant Professor, Central Department of Linguistics, T. U.  
*tara.raai@cdl.tu.edu.np*

**Diwas Rai**

M. Phil. Scholar, Central Department of Linguistics, T. U.  
*rayaday999@gmail.com*

(Received on December 27, 2025; Accepted on March 15, 2026)

### **ABSTRACT**

*This study examines the lexicostatistical and phonetic analysis of Tharu varieties spoken in Lumbini Province, Nepal. The study seeks to delineate the linguistic links and evaluate mutual intelligibility among four varieties: Dangouriya, Deukhuriya, Desouriya, and Nawalpuriya. Using a list of 210 words, data were gathered through elicitation from native speakers. The computer software COG was used to look at lexical and phonetic variation. The findings reveal that Dangouriya, Deukhuriya, and Desouriya form a core dialect cluster where phonetic similarity is from 75 to 80 percent and moderate to high lexical overlap is from 64 to 71 percent. To the contrary, Nawalpuriya emerges as a peripheral variety, showing lower lexical similarity scores from 41 to 44 percent and phonetic scores from 62 to 63 percent. This divergence suggests the influence of geographical isolation and external language contact. The pattern of phonological stability among core varieties that have lexical divergence is shaped by sociocultural factors and geographical distribution. This study contributes to identifying geographical variation and making the language policy and planning, as the Tharu language is already recommended for the provincial language in Lumbini by the Language Commission.*

**Keywords:** *Tharu language, geographical variation, lexical comparison, phonetic variation, COG software*

### **INTRODUCTION**

Tharu is one of the indigenous nationalities in Nepal. Under the National Foundation for Development of Indigenous Nationalities (NFDIN) Act, 2058 BS (2002), the Tharus are classified as a marginalized group among the five categories of the indigenous nationalities (*Adibashi Janajati*). It is noteworthy that although the Rana Tharu were originally included under the broader Tharu category, they are now officially recognized as a distinct indigenous nationality.

Tharu is an Indo-Aryan language spoken across the Tarai region of Nepal that belongs to the broader Indo-European language family (Grierson, 1903a; Eppele et al., 2012). The 2021 national census has reported that the mother tongue of Tharu is of 1,714,091 people, representing 5.88% (National Statistics Office 2021). This makes 'Tharu' the fourth largest language in Nepal. Of these, 695,949 (15.2%) reside in Lumbini Province alone, which is the largest indigenous community in the province.

## *Section Two: Sociolinguistics*

Linguistically, Tharu exhibits internal geographical diversity that has been shaped by geographical distribution. In Lumbini Province alone, there are four varieties: Dangouriya (Dang), Deukhuriya (Deukhuri), Desouriya (Banke), and Nawalpuriya (Nawalparasi West).

Early research, especially Grierson (1903a-c, 1904, 1916), included the linguistic classification of Tharu in the Linguistic Survey of India, placing it within the Indo-Aryan branch of the Indo-European family. Recent efforts, such as Epele et al. (2012) and Breton (1997) in his regional atlas, granted this classification of Tharu within the linguistic framework of South Asia.

There exist few sociolinguistic studies on the Tharu community. Boehm (1997) examined language use and preservation trends in the Nepal-India Tarai, emphasizing the challenges of bilingualism and language retention. Another sociolinguistic survey on Rana Tharu conducted by Thakur and Regmi (2012) is available. In addition to this, Leal (1971) conducted an initial phonemic analysis of the Chitwan variety, but Boehm (1998) employed a historical linguistic methodology to reconstruct the phonology of Proto-Tharu. Dhakal (2015) provided a comprehensive grammatical description of Rana Tharu. Poudel (2014) discussed the grammatical description of the Chitouniya Tharu, and thus Chaudhary (2013) presented a general overview of the Tharu language as it is spoken in Saptari District. Similarly, Dahit (2045 BS), Panjiyar Tharu (2055 BS), Choudhary (2056 BS), and Sarbahari (2067 BS) contributed to the various aspects of the Tharu language spoken in Lumbini Province.

Besides these significant works, there is a lack of the systematic, comparative analysis of geographical diversity within certain provinces. Current research exists neither focusing on the Western variants nor in the Lumbini province. So, this study, based on the previous research by Rai and Rai (2080 BS), aims to fill this gap. The objective of this study is to map the lexical and phonetic diversity among these varieties. Furthermore, this study may lay the foundation for language policy and planning in Lumbini Province, where the Language Commission has already recommended Tharu, Awadhi and Bhojpuri as provincial official languages. By analyzing geographical variations within Tharu, the research provides a linguistic foundation that can guide the effective implementation of these recommendations.

### **METHODS AND MATERIALS**

The data for this study were gathered through direct elicitation from the speakers of different Tharu varieties representing distinct geographical regions of Lumbini Province of Nepal. Precisely, the varieties were included from Dangouriya from Dang, Deukhuriya from Deukhuri, Desouriya from Banke, and Nawalpuriya from Bardaghat of Nawalparasi West. Participants were selected to represent a range of sex, age, and literacy<sup>1</sup>, embracing the sociolinguistic diversity. A total of 210 lexical items were elicited from the speakers from the respective fields<sup>2</sup>. Each elicited word was phonetically transcribed using the International Phonetic Alphabet (IPA) and organized in MS Word tables. The transcribed data were then entered into COG (Computational Comparison of Lexical Data) software to analyze the lexical items of the distinct varieties. The study explored lexical and phonetic and lexical similarity across the varieties, with analyses based on the output generated by the COG software. A 60% lexical similarity threshold was adopted as the primary criterion: varieties with less than 60% similarity were considered distinct

- 
1. Representative speakers were gathered and elicited the wordlist in the respective places during the geographical survey of the Tharu communities in Dang, Banke, Bardiya and Nawalparasi.
  2. We gratefully acknowledge the LinSuN Project, conducted under the Central Department of Linguistics, for developing the wordlist that was utilized in this study to conduct the sociolinguistic survey of Nepal.

## *Section Two: Sociolinguistics*

languages, whereas varieties with 60% or greater similarity were treated as the same language. It is noted that in certain cases, this cutoff may be applied flexibly depending on the linguistic or sociocultural considerations. This method was used to assess both the phonetic and lexical variation, comparing the Tharu varieties spoken across Lumbini Province. This study adopted the methodological insights from the previous works by Rai (2010), Rai (2019), and Rai (2023).

### **RESULTS AND DISCUSSION**

There are many different varieties of the Tharu language, each spoken in a different part of the world. These are Dangouriya, Deukhuriya, Desouriya, and Nawalpuriya. Some of these variations are somewhat understandable to each other, while others are only somewhat understandable. This study analyzes this variance through a lexical comparison that includes both lexical and phonetic metrics. The analysis examines lexical and phonetic similarities among the chosen Tharu variants to evaluate patterns of mutual intelligibility and linguistic relatedness.

#### ***Lexical Comparison***

Lexical similarity displays how many vocabulary items are shared across the chosen varieties. Higher percentages mean that the vocabulary is more shared and the degrees of mutual intelligibility are higher. Lower percentages mean that the lexicon is more different. Table 1 shows how the Deukhuriya, Dangouriya, Desouriya, and Nawalpuriya variants compare in terms of their words. The comparison shows both places where words are similar and where they are different. This gives us a better idea of how the Tharu language changes from place to place, as seen in Table 1.

*Table 1: A comparison of the words used in the several Tharu dialects spoken in Lumbini Province*

	Deukhuriya	Dangouriya	Desouriya	Nawalpuriya
Deukhuriya		68%	64%	44%
Dangouriya	68%		71%	41%
Desouriya	64%	71%		43%
Nawalpuriya	44%	41%	43%	

Table 1 illustrates the lexical comparison among Tharu varieties in Lumbini Province, highlighting patterns of both convergence and divergence in lexicon usage. Deukhuriya and Dangouriya have the most lexical similarity among the four at 68%, suggesting a considerable degree of reciprocal intelligibility in their lexical terms. Dangouriya and Desouriya have a marginally greater similarity of 71%, indicating that Dangouriya serves as a core variety connecting adjacent dialects. Conversely, Deukhuriya and Desouriya have a 64% similarity, although comparisons with Nawalpuriya generally result in the lowest percentages (41–44%). This variant exhibits the most lexical divergence among the four. These data indicate that lexical divergence exceeds phonetic divergence, potentially affected by interaction with other languages or relative geographic isolation. The data indicate a distinct pattern of core lexical similarity across central varieties (Dangouriya, Deukhuriya, and Desouriya) and increased periphery differentiation.

#### ***Phonetic variation***

Phonetic variance among Tharu varieties indicates disparities in the articulation of speech sounds across various forms. This research investigates phonetic similarity by analyzing relevant lexical words across the chosen varieties and quantifying the degree of similarity in percentage terms. Table 2 shows how the Tharu dialects spoken in Lumbini Province sound different from each other.

*Section Two: Sociolinguistics*

Table 2 *A phonetic comparison of the many Tharu dialects spoken in Lumbini Province*

	Deukhuriya	Dangouriya	Desouriya	Nawalpuriya
Deukhuriya		80%	75%	63%
Dangouriya	80%		80%	62%
Desouriya	75%	80%		63%
Nawalpuriya	63%	62%	63%	

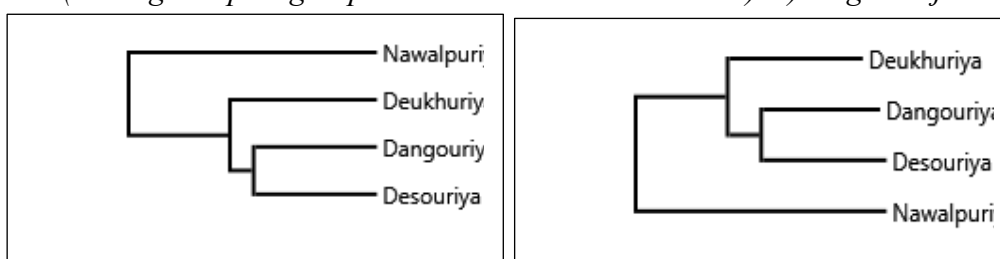
Table 2 shows the phonetic comparison of both similarity and difference among the Tharu varieties spoken in Lumbini Province. Of these, Deukhuriya and Dangouriya have the most phonetic similarities (80%). This high level of similarity is probably because the two places are close to each other and have had communication in the past. Dangouriya and Desouriya also have a high level of phonetic similarity (80%), which means that Dangouriya may be a phonetically connecting variant. The phonetic resemblance between Deukhuriya and Desouriya, on the other hand, is a little smaller (75%). It shows a moderate amount of divergence, even though they are close to one another. When comparing Nawalpuriya to the other three variants, it always has the lowest similarity ratings (62–63%). This means that it is the most phonetically different of the four kinds. The results show a pattern of phonetic diversity that goes from one end to the other. The varieties that are closer together in space or society have more phonetic features in common. On the other hand, varieties that are more isolated or on the edge, like Nawalpuriya, have phonetic qualities that are more unique.

**Hierarchical graphs**

The Tharu varieties presented in a hierarchical graph exhibit the genetic relatedness of language varieties based on the lexicostatistical similarity. There are two ways of mapping the graphs: UPGMA (unweighted pair group method with arithmetic mean) and neighbor-joining. The UPGMA is a method that helps to construct the rooted tree based on a pairwise similarity matrix. In this method, the nearest two clusters are jointly shown in a higher-level cluster. Dendrogram 1 presents a hierarchical graph showing a rooted base tree.

Dendrogram 1

a) UPGMA (unweighted pair group method with arithmetic mean) b) neighbor-joining



The network graph is a visual tool for analyzing linguistic relationships, illustrating connections in two ways: First, it reveals distinct clusters of closely related language varieties, and secondly, it maps the relative proximity between all varieties that represent degrees of similarity or difference across the entire dataset.

Figure 1 to Figure 5 present the lexical and phonetic network graph of the Tharu varieties spoken in Lumbini Province. In this visualization, varieties with high lexical similarity appear closer together, whereas varieties with low mutual similarity are positioned farther apart. This graph highlights the major varieties, Dangouriya, Deukhuriya, and Desouriya, and a closer cluster, whereas Nawalpuriya stands farther apart.

*Section Two: Sociolinguistics*

Figure 1 *Lexical network graph*

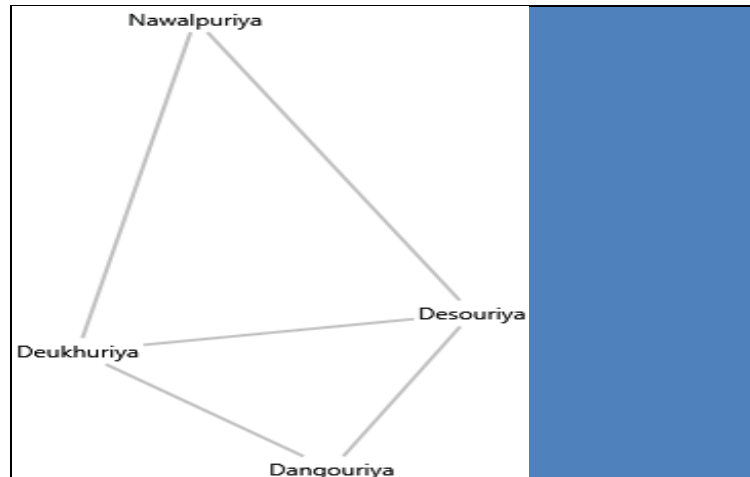
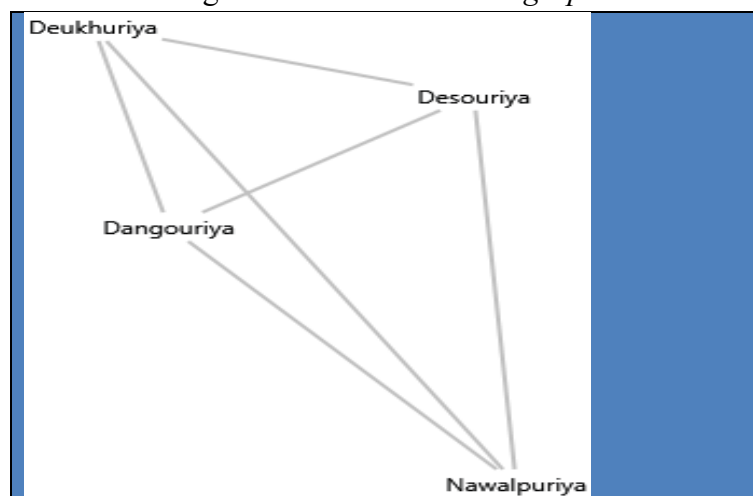


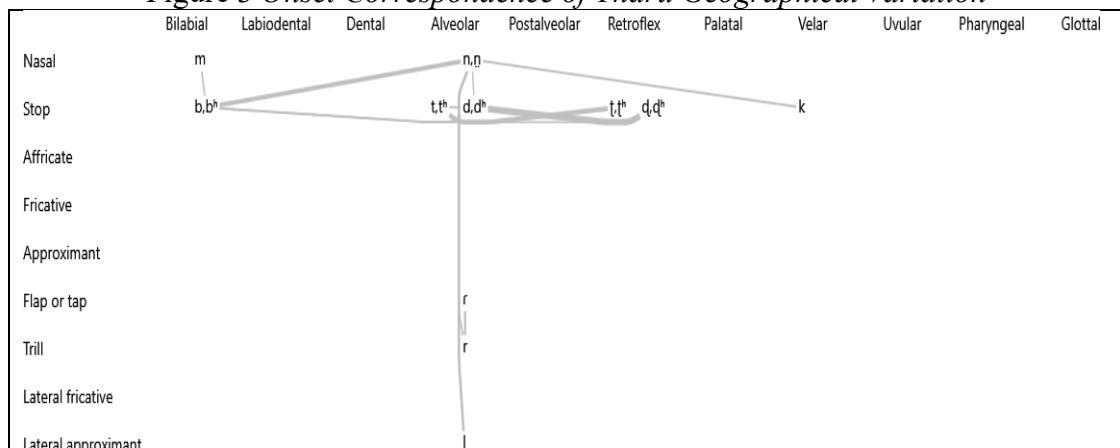
Figure 2 *Phonetic network graph*



**Global Correspondence**

Global correspondence charts are typically employed to illustrate the phonetic manifestations of consonants and vowels across many language variations. These charts illustrate the frequency thickness positions of onset, nucleus, and coda. The examination of onset correspondence emphasizes the variance in start consonants among Tharu dialects. The consonants, i.e., /p/, /t/, /k/, /m/, and /n/, are represented with broad lines, signifying limited variance. Aspirated vs unaspirated stops, such as /p<sup>h</sup>/ and /p/, or retroflex versus dental consonants, i.e., /ṭ/ and /t/, exhibit several correlations, indicating geographical diversity.

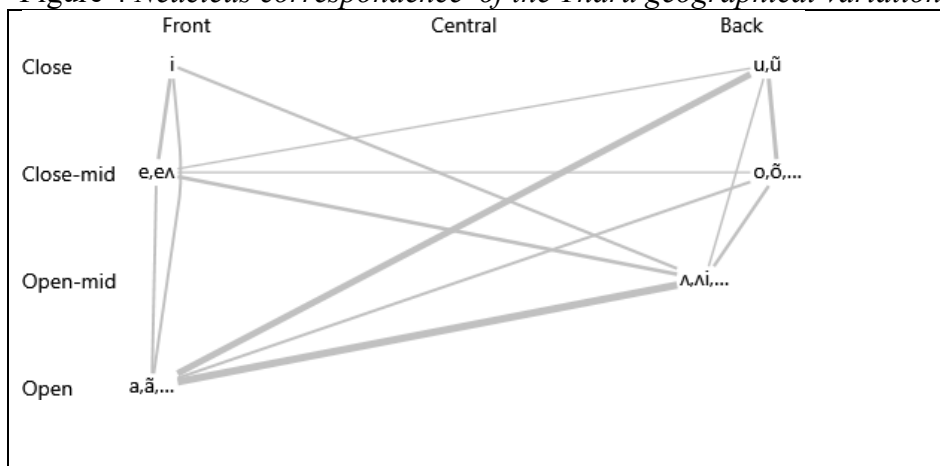
Figure 3 *Onset Correspondence of Tharu Geographical Variation*



*Section Two: Sociolinguistics*

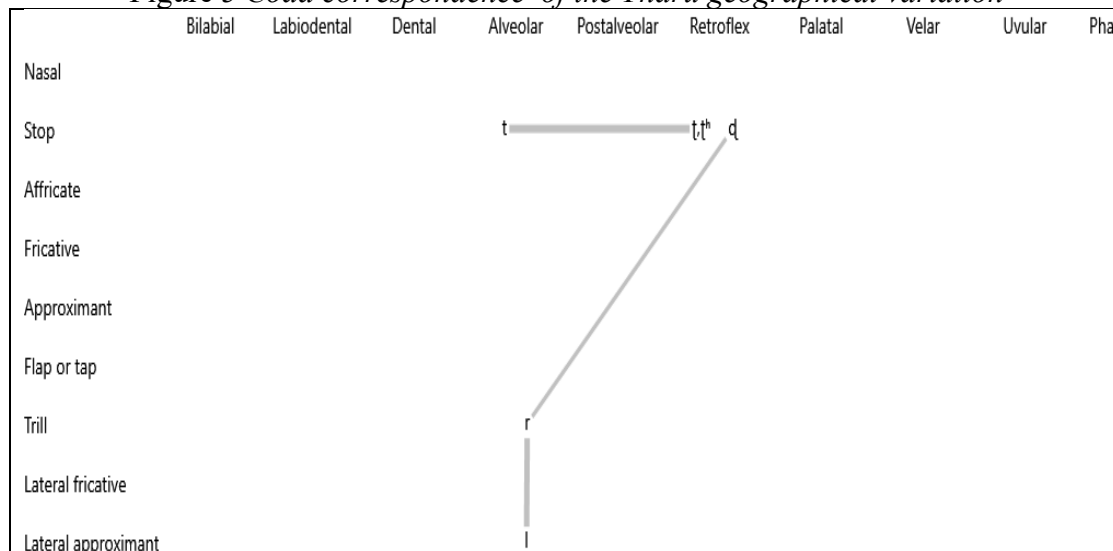
In nucleus correspondence, vowel quality and quantity across dialects are analyzed. There are two vowels merged into one variety of the Tharu that correspond to a single vowel in another, for example, /i/ and /i/ merging. Other features also lie, like nasalized /a/ and /ã/, and diphthongization, in which simple vowels correspond to diphthongs in some varieties.

Figure 4 *Neucleus correspondence of the Tharu geographical variation*



Similarly, coda correspondence exhibits variation in syllable-final consonants or consonant clusters. Common patterns include final consonant deletion, especially in casual speech; coda neutralization, where contrasts such as final /n/ and /ŋ/ are lost; and cluster simplification, where complex codas are reduced in some dialects.

Figure 5 *Coda correspondence of the Tharu geographical variation*



***Trend analysis of the Tharu varieties***

The analysis of both the lexical and phonetic variation among the Tharu varieties in Lumbini Province reveals both cohesion and divergence within the language. The varieties Denagouriya, Deukhuriya, and Desouriya show high phonetic relation that ranges from 75 to 80 percent and moderate to high lexical similarity from 64 to 71 percent. It indicates that these varieties are the close geographical variations of Tharu spoken in Lumbini Province.

There are some causes, like historical, sociolinguistic, geographical proximity, social interaction, and intermarriage, that may have brought the close relation. To the contrary, Nawalpuriya seems to be emerging as the most divergent variety, with the lowest phonetic similarity, from 62 to 63 percent, and lexical similarity, from 41 to 44 percent, compared to other varieties. This divergence may result from geographical

## *Section Two: Sociolinguistics*

isolation, reduced contact with core varieties, or external linguistic influences, such as neighboring languages or dialects. The trend analysis reveals that phonetic similarity remains generally consistent among varieties, although vocabulary exhibits greater complexity due to regional borrowing and social influences. The results show that the theories of language variation that are based on conservative phonology or phonological change are still not very strong. But language contact and mobility have had a great effect on the language than anything else. The observed patterns demonstrate the intricate interaction between geography and social networks in influencing the Tharu language landscape.

### **CONCLUDING REMARKS AND IMPLICATIONS**

In the Lumbini Province, Tharu contains four different geographical variations. This research conducted a comparative examination of the Tharu language variants in Lumbini Province, highlighting both geographical continuity and divergence. The lexicostatistical and phonetic analysis showed that Dangouriya, Deukhuriya, and Desouriya constitute a central geographical cluster with significant mutual intelligibility. The phonetic similarity is between 75 and 80 percent, and the lexical overlap is between 64 and 71 percent. Nawalpuriya, on the other hand, has a very different peripheral variety. Its lexical similarity scores are only 41 to 44 percent, and its phonetic scores are only 62 to 63 percent, which is close to or below the 60% mark that is often used to tell the difference between languages and dialects. The other characteristics, such as hierarchical graphs, network graphs, and global correspondence, likewise, support the same pattern of geographical variation for the Tharu varieties in Lumbini Province.

The results have a number of crucial effects. First, the divergence of Nawalpuriya indicates that it may be experiencing or have experienced a unique historical evolution as a different dialect or sub-language within the Tharu continuum. Second, regarding language change, the core varieties seem to maintain phonological stability, while lexical variations are found to be changing due to contact-induced factors. Finally, for language policy and preservation, this research may contribute to the urgent need for targeted documentation and revitalization strategies as the Lumbini Province is about to endorse the Language Act of the Province at the Assembly. Nawalpuriya is at greater risk of erosion or loss than those that are not closer to other varieties of the Tharu.

### **Acknowledgments**

We would like to extend our sincere gratitude to the Language Commission for providing the opportunity to conduct a lexical survey of the Tharu varieties spoken in Lumbini Province. We are grateful to Dr. Amar Giri and Mr. Amrit Yonjan Tamang, both ex-members of the Language Commission, for their support and guidance. We are grateful to the many members of the Tharu community who contributed their time and knowledge during the fieldwork. Special appreciation goes to Jayaram Adhikari (then secretary, language commission), Kamala Pandey (under secretary), Ramesh Bhattarai (section officer), Ashok Tharu from Dang, Ekaraj Choudhary from Banke, Chhabi Lal Kopila from Dang Deukhuri, Krishnaraj Sarbahari from Dang, and many others, whose insights and cooperation were essential to this study.

### **References**

- Boehm, E. D. (1998). *A phonological reconstruction of Proto-Tharu* (Doctoral dissertation). University of Texas at Arlington.
- Boehm, K. K. (1997). *Language use and maintenance among the Tharu of the Indo-Nepal Tarai* (Master's thesis). University of Texas at Arlington.
- Breton, R. J. (1997). *Atlas of the languages and ethnic communities of South Asia*. Sage Publications.

## *Section Two: Sociolinguistics*

- Central Bureau of Statistics. (2012). *National population and housing census 2011: National report*. Central Bureau of Statistics.
- Chaudhary, M. K. (2014). Language of Saptariya Tharu. *Researcher: A Research Journal of Culture and Society*, 1(2), 49–62. <https://doi.org/10.3126/researcher.v1i2.9885>
- Choudhary, M. (2056 BS). *Tharu Bhasako Lirousi Byakaran* [Lirousi grammar of Tharu language]. Byakbada Society Education.
- Dahit, G. (2045 BS). *Tharu-Nepali-English Dictionary*. Samuhik Yuva Club.
- Dhakal, D. N. (2015). *Rana Tharu grammar*. Nepal Academy.
- Epple, J. W., Lewis, M. P., Regmi, D. R., & Yadava, Y. P. (2012). *Ethnologue: Languages of Nepal*. SIL International & Linguistic Survey of Nepal (LinSuN).
- Grierson, G. A. (1903a). *Linguistic Survey of India* (Vol. V, Part I). Motilal Banarsidass.
- Grierson, G. A. (1903b). *Linguistic Survey of India* (Vol. V, Part II). Motilal Banarsidass.
- Grierson, G. A. (1903c). *Linguistic Survey of India* (Vol. IX, Part IV). Motilal Banarsidass.
- Grierson, G. A. (1904). *Linguistic Survey of India* (Vol. VI). Motilal Banarsidass.
- Grierson, G. A. (1916). *Linguistic Survey of India* (Vol. II). Motilal Banarsidass.
- Leal, D. (1971). *Chitwan Tharu phonemic summary* (Manuscript). Summer Institute of Linguistics & Institute of Nepal Studies.
- National Statistics Office (NSO). (2023). *National Population and Housing Census 2021: National report on caste/ethnicity, language, and religion*. Government of Nepal.
- Panjiyar Tharu, T. (2055 BS). *Tharu Bhasa: Ek Parichay* [Tharu language: An introduction]. Bouli Choudhary.
- Poudyal, K.P. (2014). *A Grammar of Chitoniya Tharu*. LINCOM EUROPA.
- Rai, T. M., & Rai, D. (2080 BS). *Lumbini pradesmā boline Thāru bhāṣāko bhaugolik bhedharu* [Geographical variation among the Tharu spoken in Lumbini Province]. Language Commission.
- Rai, T.M. (2010). Lexical similarities in Kirati languages: evidence from Koyee, Dumi, and Sampang. *Isilim*. 4. 80-84.
- Rai, T.M. (2019). Lexical comparison in Hayu. A lexicostatistical analysis. *Gipan* 4. 2019. 162-174.
- Rai, T.M. (2023). Tamu (Gurung), Tamang, Thakali, Chhantyal Bhasabichko Antarsambandha: Kosiya Tathyankaka Adharma Tulana [Interrelation among the Tamu (Gurung), Tamang, Thakali, and Chhantyal languages]: Lexical comparison. *Bhasalok*. 3.57-75.
- Regmi, D. R., Thakur, G., & Regmi, A. (2012). *A sociolinguistic survey of Rana Tharu (as spoken in Kailali and Kanchanpur districts, Far-Western Development Region, Nepal)*. Report. Linguistic Survey of Nepal, Central Department of Linguistics, Tribhuvan University.
- Sarbhari, K. (2067 BS). Tharu Bhasa Sahityako Ruprekha [An outline of the Tharu language literature]. *Sayapatri*. 21. 44-45.