

Building a corpus of Vaiphei, a Tibeto-Burman language of Northeast India

Northeast India constitutes a region of high cultural and linguistic diversity with many yet un- or under documented languages (cf. Post et al. 2015). Embedded in a project on person marking in South Central (a.k.a. “Kuki-Chin”) Tibeto-Burman languages, the author intends to start a documentation of the Vaiphei language in order to contribute to language documentation efforts in this region. Vaiphei is a language of the South Central subbranch of Tibeto-Burman spoken by several 10.000 speakers. Most speakers reside in the state of Manipur with smaller communities existing also in the states of Assam and Meghalaya, among others (cf. Singha 2024: 1). While a number of descriptive studies exist (Suantak 2013; Prichard 2018; Murakami 2018), so far no extensive language documentation has been undertaken. Corpus data are therefore lacking. Furthermore, the existing studies focus on the variety/ies spoken in Manipur to the exclusion of the varieties spoken in other Northeast Indian states.

This contribution attempts to discuss how to approach the process of corpus building for Vaiphei from a mostly methodological perspective. As part of their PhD project, which is supposed to produce a grammar sketch and corpus-based study, the author intends to conduct fieldwork in the Vaiphei villages of Assam, where a different variety of the language is spoken compared to Manipur. While the earlier phases of grammatical analysis will require extensive elicitation, a major emphasis shall be put on creating a rich corpus of spoken interaction data in close cooperation with the Vaiphei-speaking community. The goal is to move beyond the elicitation-heavy paradigm that has characterized much of the linguistic work in Northeast India. By the same token, the collection of interaction data is intended to be undertaken from the outset, rather than be postponed to a later stage of fieldwork. Naturally, community involvement is central: speakers will be invited to contribute not just as linguistic consultants but also as transcribers and potentially documenters as to ensure a more grounded and participatory documentation process. The data will be collected and archived following the FAIR and CARE principles (cf. Carroll et al. 2021). As a further addition to the slowly growing number of corpora of under documented languages spoken around the world, it is hoped that the Vaiphei data will be useful not only for the purpose of grammar writing as envisioned for the authors PhD thesis but also be of use for linguists investigating other phenomena in the future. On the methodological and technical level, a major challenge will potentially lie in capturing the dynamics of spontaneous multilingual speech, as well as dealing with typical features of spoken interaction data, such as overlaps. Multimodality could also potentially be interesting for the intended research project, even though at this stage of the project its importance cannot be adequately assessed.

Keywords: Vaiphei, Tibeto-Burman, language documentation, corpus building, fieldwork, Northeast India

References

- Carroll, Stephanie Russo & Herczog, Edit & Hudson, Maui & Russell, Keith & Stall, Shelley. 2021. Operationalizing the CARE and FAIR Principles for Indigenous data futures. *Scientific Data*. Nature Publishing Group 8. 108. (doi:10.1038/s41597-021-00892-0)
- Murakami, Takenori. 2018. Two Comparative Forms in Vaiphei. *Proceedings of the 51st International Conference on Sino-Tibetan Languages and Linguistics*, 711–722. Kyoto: 第 51 回国際漢蔵語学会実行委員会・京都大学白眉センター.
- Post, Mark W. & Morey, Stephen & DeLancey, Scott (eds.). 2015. *Language and Culture in Northeast India and Beyond: In honor of Robbins Burling* (Asia-Pacific Linguistics 23). Canberra: Asia-Pacific Linguistics.

- Prichard, Jesse Andrew. 2018. *Verbal Stem Alternation in Vaiphei*. Portland: Portland State University. (MA thesis.)
- Singha, Kabita. 2024. A Sociolinguistic Exploration of the Vaiphei Tribe. *International Journal for Multidisciplinary Research (IJFMR)* 6(4). 1–4.
(doi:10.2139/ssrn.5054029)
- Suantak, Khawlsongkim. 2013. *Vaiphei phonetics phonology and morphology: a descriptive study*. Shillong: North-Eastern Hill University. (PhD thesis.)