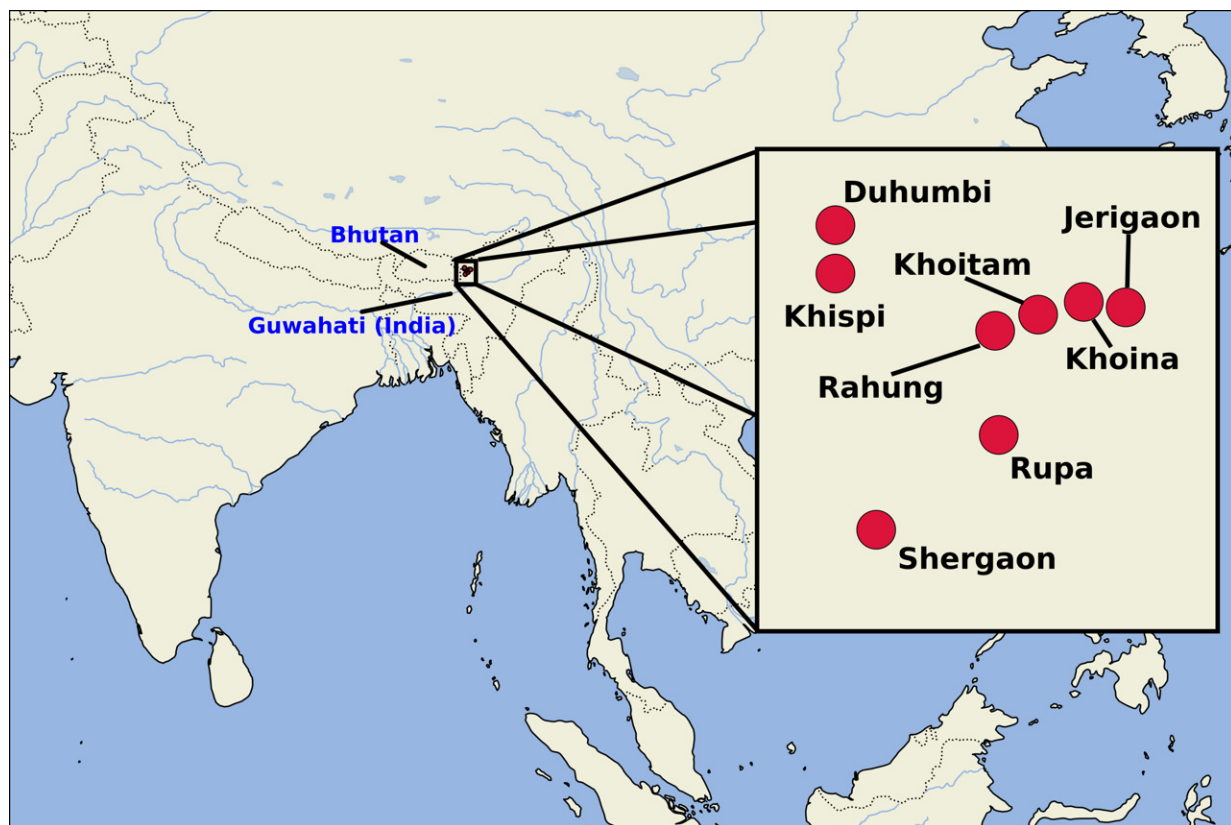


Linguists predict unknown words using language comparison

April 27 2021



Map showing geographic locations for speakers of eight Western Kho-Bwa languages. Credit: Johann-Mattis List

For a long time, historical linguists have been using the comparative method to reconstruct earlier states of languages that are not attested in

written sources. The method consists of the detailed comparison of words in the related descendant languages and allows linguists to infer the ancient pronunciation of words which were never recorded in any form in great detail. That the method can also be used to infer how an undocumented word in a certain language would sound, provided that at least some information on that language, as well as information on related languages is available, has been known for a long time, but so far never explicitly tested.

Two researchers from SOAS University of London and the Max Planck Institute for the Science of Human History have recently published a paper in the renowned international journal for [historical linguistics](#), *Diachronica*. In the article, they describe the results of an experiment in which they applied the traditional comparative method to explicitly predict the pronunciation of words in eight Western Kho-Bwa linguistic varieties spoken in India. Belonging to the Trans-Himalayan family (also known as Sino-Tibetan and Tibeto-Burman [language](#) family), these varieties have not yet been described in much detail and many words had not yet been documented in [field work](#). The scholars started their experiment with an existing etymological dataset of Western Kho-Bwa varieties that was collected during fieldwork in the Indian state of Arunachal Pradesh between 2012 and 2017. Within the dataset, the authors observed multiple gaps in which the word forms for certain concepts were missing.

"When conducting fieldwork, it is inevitable that you miss out on some words. It's kind of annoying when you observe that afterwards, but in this case, we realized that this was the perfect opportunity to test how well the methods for linguistic reconstruction actually work," says Tim Bodt, first author of the study.

The researchers set up a computer-assisted workflow to predict the missing word forms. The classical methods are traditionally applied

manually, but the new computational solutions helped the scholars to increase the efficiency and reliability of the process, and all results were later manually checked and refined. To increase the transparency and validity of the experiment, they then registered their predictions online.

"Registration is incredibly important in many [scientific fields](#) because it ensures that researchers adhere to good scientific practice, but as far as we know it has never been done in historical linguistics," says Johann-Mattis List, who carried out the computational analyses of the study.

"By registering our predictions online, we made sure we could no longer modify our predictions in light of the results we obtained during our subsequent verification process," Bodt, adds.

With predictions in hand, Bodt then traveled to India to verify the predicted words with native speakers of the Western Kho-Bwa languages. After asking the participating local language consultants to provide their words for the concepts under investigation, the authors compared these attested words with their earlier predictions. Based on the proportion of correctly predicted sounds per word form, the predictions were correct in 76% of all cases, which is remarkable given the limited amount of information that was used to predict the word forms. Moreover, the scholars were able to identify several reasons why certain predictions did not match the actual attested forms in the languages.

"The more we know about a language family in general, the better we can predict unknown word forms. This is all possible, because languages change their sound systems in a surprisingly regular manner," says List. "Despite the fact that so little was known about the Western Kho-Bwa languages and their linguistic history, we were able to show through our experiment that regular sound changes result in predictable word forms. In turn, our experiment has greatly improved our understanding of the

Western Kho-Bwa languages and their linguistic history."

Apart from giving a concrete example for the power of the methodology of historical linguistics and the value of their experiment for linguistic studies, the authors identify certain additional benefits of predicting words in linguistic research.

"Predicting words increases the transparency and efficiency of our research and our fieldwork. This is crucial in light of rapid language loss and limited funding for descriptive linguistic work. Moreover, it also has an educational aspect since it encourages speakers to reflect on their own linguistic heritage," says Bodt.

The researchers hope that the results of their ground-breaking experiment will encourage other linguistic field workers, descriptive linguists, and historical linguists to follow suit, and make more explicit and conscious use of the regularity of sound change and predictions of word forms.

More information: Timotheus A. Bodt et al, Reflex prediction, *Diachronica* (2021). [DOI: 10.1075/dia.20009.bod](https://doi.org/10.1075/dia.20009.bod)

Provided by Max Planck Society

Citation: Linguists predict unknown words using language comparison (2021, April 27) retrieved 25 June 2024 from

<https://phys.org/news/2021-04-linguists-unknown-words-language-comparison.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.