

# A database could help revive the Arapaho language before its last speakers are gone

Andrew Cowell, Professor of Linguistics, University of Colorado Boulder

Published: November 26, 2025 7:17pm EDT



There are fewer than 100 speakers of the Arapaho language today.

*Mark Makela/Getty Images*

I was hired at the University of Colorado Boulder in 1995 as a language professor. I relocated from Hawaii, where I had learned the Hawaiian language.

When I arrived in Colorado, I decided I needed to learn about the Indigenous language of the Boulder and Denver area, Arapaho. The Arapaho people had occupied the area for many years until they were forced to leave in the 1860s.

I first visited the Northern Arapaho people on the Wind River Reservation in Wyoming in 1999. At that time, there were hundreds of speakers of the Arapaho language.

Today, there are less than 100, and all are over the age of 70.

The Arapaho people in Wyoming and Colorado believe their language can still survive, and so do I. That's why I am working to combine decades of language documentation with new technological approaches in order to help revive the language.

## **Loss of Native languages**

Many Native American languages currently have few Native speakers, and the speakers are typically the oldest members of the community. The languages of the Wichita and Kansa people, for example, are among many that are no longer spoken at all.

Native American languages have been in decline in the face of Euro-American pressure for centuries.

On the Great Plains, this decline accelerated after World War II when Native soldiers came home after seeing prosperity off the reservation.

Arapaho elders tell me that bilingual parents decided to speak only English to their children to improve their chances of success in life. They were certain the tribal languages would come “later.”

But “later” didn’t happen. Boarding schools had already been suppressing the language, and now economic improvements brought cars, radios and televisions to Wind River, further promoting the use of English. Without language exposure in the home, children were not able to acquire good speaking abilities.

Today, however, tribal communities around the country increasingly want to maintain or reacquire their languages. Efforts to do this have been going on for several decades, with some successes, such as the Mohawk language of New York and Canada, Cherokee in Oklahoma and North Carolina and the Blackfoot language of northern Montana.

In most places however, numbers of Native speakers continue to decline, while learning among younger speakers progresses slowly.

## **Uses of data for curriculum**

My early work focused on documenting the Arapaho language. Past linguists working with Native languages typically focused on traditional storytelling, as well as audio-recorded data. But my interest in anthropology led me to focus on conversation and everyday interaction. I also recorded on video to capture social settings, gestures and sign language. And to better understand the role of the language in daily use, I worked to become a good speaker myself.

I have compiled my documentation into a database that contains over 100,000 sentences of natural Arapaho speech. All of this has been transcribed, translated into English and accompanied by detailed linguistic analysis.

The database is further supported by an online learning site and an online dictionary of around 25,000 entries. They are among the largest such resources for an Indigenous language, though resources do exist for other languages, such as Yurok.

*Courtesy of Andrew Cowell. 13.8 KB (download)*

## From documentation to curriculum

In response to the Arapaho people's goal of language revitalization, my own work has shifted from documentation to assisting teachers, students and curriculum developers. The database turns out to have great value in this area.

Adult learners can watch the videos along with the Arapaho transcriptions or English translations, or both, and review the detailed grammatical analysis.

However, it is quite difficult for young learners to immediately benefit from listening to natural discourse. That's why carefully graded curricula are crucial. Unlike for commonly taught languages such as French or Spanish, materials for most Native American languages are just being developed.

Arapaho can be challenging to learn because its structure is quite different from English. Many small chunks of meaning are combined to produce long, complex words. For example, an English speaker can start with "happy" and produce "un-happi-ness." Arapaho speakers typically add three, four or even five prefixes, and multiple suffixes as well. A speaker can say the word "niibeetwonwoteekoohunoo" – which has six separate meaningful chunks. This translates to English, "I want to go and drive to town."

There is little value in memorizing such complex words, just as English learners don't memorize entire sentences. Instead, Arapaho learners need to understand the separate parts, and how they combine.

Previous efforts have succeeded in teaching children to speak basic Arapaho. The challenge now is to keep improving their Arapaho language abilities, using a graded curriculum that continues through all school levels.

The database can identify and label the individual chunks of words, and assign meanings to each chunk. A beginner's dictionary of 1,300 entries has been created by calculating the overall frequency of base words in the 100,000 sentences, and then selecting only the most common ones.

The list has been broken down further to produce target vocabulary for each grade level. Smaller chunks of prefixes and suffixes are also measured, and sequential grammar-learning goals can be produced based on frequency and complexity.

A draft Arapaho learning sequence has been created, with 44 stages. It is now possible for the first time to produce a full, progressive language curriculum for Arapaho. The next step is to develop more curricular materials and train teachers to use them.

The sequence of 44 stages is now being introduced at Wyoming Indian Elementary School, the first school on the Wind River Reservation to pioneer dual-language classrooms.

## Limitations of technology

Technology is not a magic bullet, however. Only Native people can save their languages, by choosing to learn and speak them.

Because artificial intelligence works using large language models, it needs billions of words of discourse to be trained effectively in a language. No Indigenous language has nearly that amount of data, so the capacity of AI to address Native language endangerment is limited. Moreover, many Indigenous communities are wary of AI due to concerns over data sovereignty and cultural property rights.



The author, Andrew Cowell, is recognized for his Arapaho language revitalization at a 2018 ceremony on the Wind River Reservation in Wyoming.

*Courtesy of Andrew Cowell.*

My own old-fashioned experience as a learner and teacher has proved crucial. I can see where difficulties lie for learners, and how to fine-tune computational measurements and predictions. I've learned that success in helping revitalize Native languages depends on researchers building long-term relationships with Native peoples and, ideally, speaking Native languages. Only then can new technologies be applied most productively.

*Andrew Cowell currently receives funding from National Science Foundation. Past funding related to the work described here has come from the American Council of Learned Societies and Hans Rausing Endangered Language Documentation Programme. He has received compensation from elements of the Northern Arapaho Tribe and the Southern Cheyenne and Arapaho Tribe for some of his assistance and consultation.*

This article is republished from The Conversation under a Creative Commons license.