[**IMPACT ON OUR STATE**](https://www.unc.edu/discover-theme/state-impact/)

The sanctity of Cherokee

As a result of systemic oppression, there are fewer than 200 native Cherokee speakers in North Carolina. To keep the language alive and pass it to the next generation, UNC-Chapel Hill researcher and Eastern Band Cherokeean citizen Benjamin Frey has teamed up with computer scientists Mohit Bansal and Shiyue Zhang to create a new translation model and grow the literary library of works available in Cherokee.

Story by Elise Mahon, UNC Research. Video by John Roberts, University Communications, Monday, November 15th, 2021

* Play Video
* [Share on Facebook](https://www.facebook.com/sharer/sharer.php?u=https://www.unc.edu/discover/the-sanctity-of-cherokee/)
* [Share on Twitter](https://twitter.com/home?status=The%20sanctity%20of%20Cherokee%20https://www.unc.edu/discover/the-sanctity-of-cherokee/)
* [Share on Pinterest](https://pinterest.com/pin/create/button/?url=https://www.unc.edu/discover/the-sanctity-of-cherokee/&media=https://www.unc.edu/wp-content/uploads/2021/11/CLR-4.jpg&description=The%20sanctity%20of%20Cherokee)
* Share via Email
* [Print Friendly](https://www.unc.edu/discover/the-sanctity-of-cherokee/)

**Press the play button above to watch the video**

Imagine if the language you speak at home to communicate with your family and friends — the language you think in — was prohibited from use at school. This is not a problem that native English speakers in America ever face. They simply think, open their mouths, and speak, answer a question, share their thoughts, and communicate with their peers, most of whom also primarily speak English. But language hasn’t always been so singular.

Before the arrival of Europeans, around [300 distinct](http://mithun.faculty.linguistics.ucsb.edu/pdfs/1998%20Nat%20Am%20Lgs%20Encarta.pdf) mutually unintelligible languages were spoken in North America. Now only 187 remain.

“Cherokee is a language that I didn’t inherit from my mother because of the violence that my grandmother was subjected to at federal boarding schools,” says [Benjamin Frey](https://americanstudies.unc.edu/ben-frey/), assistant professor in the Department of American Studies. “She was beaten for speaking the language and had her mouth washed out with soap, so she didn’t pass the language on because she didn’t want anybody else to be treated that way.”

While there are upwards of 2,000 speakers globally, most in Oklahoma, there are fewer than 200 native Cherokee speakers left locally. And Frey, a professor in the American studies department, wants to change that.

Most recently, he teamed up with computer science researchers [Shiyue Zhang](https://www.cs.unc.edu/~shiyue/) and [Mohit Bansal](https://www.cs.unc.edu/~mbansal/) to create a [new translation tool](https://chren.cs.unc.edu/)similar to Google Translate. They hope the tool is eventually accurate enough to translate entire novels, providing children in immersion schools with popular children’s literature, like the “Harry Potter” series, in Cherokee.

“One of the things I’ve seen is that people will frequently ask for translations, and speakers are generally happy to give them. But larger translation projects really take up a lot of time and energy for our speakers who are mostly elderly,” Frey says. “It was one of those situations where I thought, ‘Well, this is all tremendously important and amazing work. Wouldn’t it be more efficient and beneficial if we could do it faster?*‘”*

**A data-hungry demo**

When Frey first approached Zhang and Bansal about building a computer translation tool, they hopped on the opportunity immediately. Frey spent countless hours working to enter 11,000 lines of Cherokee-English parallel sentences — something he was quite proud of — but when he shared that with Zhang and Bansal, he learned that wasn’t very much data compared to other languages.

The computer scientists grew curious. Were there other resources they could tap into? They asked what languages Cherokee was related to, hoping to find parallels they could use to train their new model. Frey started to list off related Native American languages: Seneca, Cayuga, Onondaga, Oneida, Mohawk, Tuscarora. But neither Zhang nor Bansal had ever worked with them before. They asked if any of these languages had robust corpuses or even machine translation tools of their own.

“I was like, ‘No, because of the genocide. People have been trying to kill them for 300 years.’ And so we had to sort of recalibrate expectations,” Frey explains.

With Mohit serving as the project’s principal investigator and Zhang’s advisor, the team had to augment the volume of data they had, resorting to experimental methods that would help to train the translation tool to improve by using automatic data processing and various machine-learning techniques.

“Current models are very data-hungry,” says Zhang, a fourth-year doctoral student.

Since translation models have successfully been produced to translate more common languages — like German, Chinese, and Russian — to English, the researchers were able to use those models to train their new tool. Zhang shares another important method they’re using to “teach” the translation tool, called a human-in-the-loop model. This is a feedback box for native speakers to provide suggestions and corrections to translations that can be plugged back into the model to improve its accuracy.

Zhang hopes the tool — which is still in the early stages of development and not yet ready for public use — can help provide academic literature in Cherokee. The tool is one of several projects where members of the College of Arts & Sciences’ computer science department use their skills to support researchers in other fields.

“Most research papers are in English by default,” she says. “Having these resources available in endangered languages can make the research community more inclusive and diverse.”

**Doing the language justice**

Like most Americans, you’ve probably used Google Translate at least once. Maybe it was to check a vocabulary word from a long-forgotten French class or to look up the translation of a sign you passed on the street. Or maybe you used it to “double-check” the answers for your Spanish homework, only to find the next day in class that the answers you translated were just slightly off in syntax or conjugation.

Since languages develop to meet the needs of the people speaking them and every culture has different values and realities to communicate, it’s logical that there won’t always be direct and complete translations for some concepts. Frey believes finding a way to interpret and grasp some of these concepts is vital.

“We don’t slice up the color blue the same way the Russians do, and so if we don’t have words, for things that are sliced up that different way, then it’s more difficult for us to understand those fine distinctions,” he says. “Being able to preserve languages that have lots of different ways of slicing up the world epistemologically is important for human culture, human heritage, and potentially problem-solving, too. If you want a solution to a problem that you’re just at your wits end on, maybe go to somebody who fundamentally thinks differently than you do, who thinks outside the box.”

The structure of the language we speak and think in shapes the way we see the world — a [concept developed by linguist Edward Sapir](https://www.linguisticsociety.org/resource/language-and-thought) in 1929. Languages contain words that encapsulate different slices of reality experienced by different cultures, making them encoded with cultural knowledge. They reflect the values of a culture and the perspective they have on the world. In Cherokee, every word inherently references relationships since, just like the people that speak it, nothing is truly isolated.

“The only way I’m going to get close to [saying] the full picture is to be in relationship with people, so I *have* to be in relationship with everybody in order to get the whole picture of what we as human beings understand,” Frey explains.

Frey believes most people miss the “justice angle” when thinking about endangered languages, focusing too much on the benefits these languages could bring to them rather than acknowledging their inherent value and sanctity.

“People are sacred just as they are. They don’t need to think like you do or be like you to be acceptable,” Frey says. “These people never should have been taken away from their homes and beaten for speaking their language. That was a massive wrong perpetuated against indigenous people. [Cherokee] preserves our identity as a unique people and connects us to our personas, our heritage, our ancestors.”

[**Learn more about the Cherokee language at Endeavors.UNC.edu**](https://endeavors.unc.edu/the-sanctity-of-cherokee/)