

Researchers use AI to help translate the Bible

July 10, 2023

WASHINGTON (RNS)—Out of the 7,100 languages that exist, the Bible has been translated into more than 700, making it the most-translated book in the world. Yet, those remaining languages—many of them extremely rare—have vexed Bible translators for decades.

Two scientists are looking to new advancements in artificial intelligence to help close the gap.

“We want to reach all the languages on Earth. The goal is to reach everyone,” said Joel Mathew, a research engineer who alongside Ulf Hermjakob recently launched the Greek Room, an AI-powered technology to help streamline the highly technical process of biblical translation.

Combining Hermjakob’s long experience with natural language processing technologies and Mathew’s field knowledge of Bible translation, the two USC Information Sciences Institute researchers developed the technology with an aim to target “very low-resource languages that are not even in the top 500,” Mathew said.

The Greek Room includes three main tools: spell-checking, world alignment that ensures consistency in translation, and Wildebeest, used to detect improper characters in a script.

The two scientists met in 2015 when Mathew joined USC to complete a master’s degree in computer science. There, he met Hermjakob in the AI division of the Information Sciences Institute. They bonded over a shared passion for languages and their Christian faith.

Mathew, the son of two Bible translators, has observed firsthand the difficulties that come with manual translation by local church members. In his hometown, New Delhi, he took notes of all the tasks that technology could accomplish.

Spell-checking usually requires many people and time, he explained. In the context of translation into rare languages, only local church members are qualified, and they don't have technology to back their work.

"These are not trivial problems. These are very hard problems. But big companies are not interested in solving them. It's not their business model to target very rare languages," he said.

When Mathew shared with him some of the problems Indian translators faced on the ground, Hermjakob jumped at the occasion.

"I always had this feeling to know how, at some point, I could apply my skills to my faith," said Hermjakob, who earned a Ph.D. in computer science from the University of Texas.

Change the traditional model of biblical translation

With their project, Mathew and Hermjakob want to work on languages that do not even have a written system, grammar codes, dictionaries or spell-checkers.

"We are thinking of languages like Uyghur or Oromo," Hermjakob said. Oromo is spoken in Ethiopia and Northern Kenya.

Recently, they have been approached by an Indian consultant specifically interested in the spell-checking and word-alignment tool for Bible translation in Kolami, a language spoken in western India that counts

130,000 native speakers.

The Greek Room also hopes to change the traditional model of Bible translation. Historically, translations were done by Western missionaries, who could only work on two languages at most in their lifetime, Hermjakob explained. With the Greek Room, the two researchers encourage a local church-driven model.

“Local churches and local language communities are asking for translations of the Bible in their heart language,” Mathew said. In a multilingual context, he added, the heart language is the one in which people express their deepest feelings and is usually their native language.

This first version of the Greek Room focuses on quality control so translators can prioritize other tasks that require more judgment, like finding a way to translate a concept that doesn’t exist in a given language. In their next version, the two researchers want the tool to suggest better translations.

Now that their codes and data are available on GitHub, they hope other users will integrate their research into their tools and innovate further.

Their initiative, supported by the Wycliffe Bible Translators USA organization, is part of a broader program directed by Every Tribe, Every Nation that hopes to make the Scripture available in every language by 2033.