Hindi to Gondi Translation Tool

August 18, 2020
The app, developed by Microsoft Research Lab, CGNet Swara and IIIT Naya Raipur, during the lockdown, hopes to motivate youth from the Gond Adivasi community to learn the Gondi language. The **Interactive Neural Machine Translation (INMT) tool translates sentences from Hindi to Gondi and vice-versa.**

Hindi to Gondi Translation Tool

This initiative is being led by Microsoft Research Lab and CGNet Swara, an Indian voice-based online portal that gives people in the forests of central tribal India a platform for expression by reporting local news and stories through a phone call. Interestingly, most of the project has been executed during the covid-19 induced lockdown. For the past four months, nearly 150-plus **Gondi speakers, spread across the six states of Maharashtra, Chhattisgarh, Odisha, Andhra Pradesh, Telangana and Madhya Pradesh**, from all walks of life have been sitting at home, translating sentences from Hindi to Gondi and back.

Even though the language is spoken by nearly 12 million Gond Adivasis, **it is not standardised, with different versions spoken in the six states. There is also no written literature in the language, with the dialects having been passed down orally over the centuries. And hence there are no local teachers**, and the ones coming from outside only spoke Hindi. As a result, the new generation speaks mostly Hindi and the language is disappearing. This tool is an effort to save it. This project is an **extension of the Microsoft Research Lab's work on natural language processing, as part of which it focuses on low-resource languages such as Gondi, where so little data is available.** The team took on this project after

getting to know of CGNet Swara's various other language initiatives.

Gondi is a very good language to use as a case study as it has a substantial speaker base across six states. It is not endangered and yet zero resources are available for the same. Through CGNet Swara, the Microsoft Research Lab became aware of the various issues that the Gond Adivasis face, and how access to the language could help the cultural identity of the community. By bringing together technology and language and providing easy access to people, the team hopes to inspire others, who want to do similar work with other communities.

Instead of a top-down approach, in which the decisions of the technologists would be imposed on the community, **the project focused on ideas and desires of the community members.** The workshop in Bengaluru, attended by academic partners, the CGNet Swara team and community members, turned out to be an enriching experience for all present. However, an Interactive Neural Machine Translation tool requires a substantial amount of data for the model to work. And there was very little data available in Gondi at that time.

As data started coming in, the project managed to overcome its first hurdle last month by crossing a bank of 20,000 sentences. Today, the base has increased to 35,000 sentences and the aim is to have translations of at least 1 lakh sentences. With the app, a window will open up for the community. Anything written in any major language in the world can then be translated and communicated to them. **By making it a mobile app and not a browser-based one, the idea is to make it accessible in places where not many devices are available and with inconsistent internet services.** It can even work offline. This is one of the many language projects that CGNet Swara has been working on, **one of the foremost being the standardised Gondi dictionary.** Already 3,000-plus words have been added to it. Recently the team has worked on another project, **translating 400 children's books by Pratham Books** in

Gondi.