Project ELLORA

February 2, 2023

In news-Microsoft's Project ELLORA is helping small languages like Gondi, Mundari become eloquent for the digital world.

About the Project ELLORA (Enabling Low Resource Languages) in India-

- To bring 'rare' Indian languages online, Microsoft launched project ELLORA or Enabling Low Resource Languages in 2015.
- Under the project, researchers are building digital resources of the languages.
- They say that their purpose is to preserve a language for posterity so that users of these languages "can participate and interact in the digital world."
- **The main goal of ELLORA is to impact underserved communities through enabling language technology by creating economic opportunities, b**uilding technological skills, enhancing education and preserving local language and cultures for future generations**. ELLORA aims to do this by:**
    - **Data:** New/Innovative methodologies for data design and collection, e.g., gamification of data collection, crowdsourcing.
    - **Language Technology Systems:** Designing new techniques and framework/architecture for technology for low resource languages, building Speech and NLP systems for low resource languages.
    - **Applications:** At scale deployments of language technology applications that impact the community.
- **Microsoft Research (MSR) has chosen to focus on three of these for now.**

Gondi language-

- It is a South-Central Dravidian language, **spoken by**

about three million Gondi people, chiefly in the Indian states of Madhya Pradesh, Maharashtra, Chhattisgarh, Andhra Pradesh, Telangana and by small minorities in neighbouring states.

- Gondi is a unique script, which is **perhaps the only script in the country besides Urdu which is written right to left, also has three or four versions.**
- Although it is the language of the Gond people, it is highly endangered, with only one fifth of Gonds speaking the language.
- Another unique quality of the script is that in the northern and central parts of India, it is the only language, barring Gujarati, which has a script of its own.
- All other north and central Indian languages use the Devnagri script.
- Gondi has a rich folk literature, examples of which are marriage songs and narrations. Gondi people are ethnically related to the Telugus.

Mundari language-

- Mundari is a **Munda language of the Austroasiatic language family** spoken by the Munda tribes in eastern Indian states of Jharkhand, Odisha and West Bengal.
- It is **closely related to Santali. Mundari Bani, a script specifically to write Mundari**, was invented by Rohidas Singh Nag.
- It has also been written in the Devanagari, Odia, Bengali, and Latin writing systems.

Idu Mishmi language-

- It is a **small language spoken by the Mishmi people in Dibang Valley district**, Lower Dibang Valley district, Lohit district, East Siang district, Upper Siang district of Arunachal Pradesh and in **Zayü County of the Tibet Autonomous Region, China.**

- It is considered an endangered language.
- The Idu Mishmi people did not usually have a script of their own. When needed Idu Mishmis tended to use the Tibetan script.
- **Currently the Idu Mishmi have developed a script known as "Idu Azobra".**