

Interview

With **Andras Kornai**, Professor, the Computer and Automation Research Institute, Hungarian Academy of Sciences, Budapest, by **Pandurang Hegde**, activist Chipko/Appiko movement.

Professor Kornai's paper on 'Digital Language Death'¹ provides information on the reach and presence of languages in the digital world and its impact on the survival of languages in the coming years. Of approximately 7000 languages spoken today, some 2500 are generally considered endangered. After extensive research he argues that this consensus figure vastly underestimates the dangers of digital language death; less than five per cent of all the languages can still ascend to the digital realm, but most languages are destined towards a massive die-off caused by the digital divide.

Andras Kornai has co-authored a position paper on the Indian Subcontinent Language Vitalization project with the objective of turning as many languages and dialects of the subcontinent as feasible into digitally viable languages. With the strong multilingual base, he feels that there is scope for halting the human cost of digital language death by helping digitally still languages to access the digital realm through 'a champion' language like Hindi or other regional languages other than English.

Pandurang Hegde: *How do you look at the major challenges and prospects of languages in the digital era? Do you think that the digital realm provides equal space to all languages?*

Andras Kornai: No, of course not. Languages whose speakers are more literate, more educated, and are concentrated in areas with better internet access have a huge advantage. The main challenge, at least the way I see it for India, is to make sure that intellectuals don't abandon their native language in favour of English,

Hindi, or other well-entrenched regional languages, but they build up education, literature, news, and even business in their mother tongues.

The digital era has posed a civilization challenge to the diversity of languages, as your research has shown; only a few languages have the chance of survival. Your findings show that only five percent of the languages that exist today can ascend in the digital era, and the other 95 percent of languages will be pushed towards extinction in the coming years. What are the main reasons behind this?

The key reason is the lack of overlap between the digitally literate and the native speaker populations for these languages. If, as the famous saying goes, it takes a village to raise a child, it takes a town, with a university, to raise a community of intellectuals. It is not sufficient to send a smart kid to some faraway college – what we need is a community of such kids who will want to talk to each other by texting in their native language.

Do you think that the extinction of minority languages is a natural phenomenon of an evolutionary process in which dominant languages appropriate the space available to other minority languages in society? Can one argue that the present domination of a few languages in the digital world is part of this evolutionary process, or is it something entirely different?

This is new, and qualitatively different from the wellknown processes whereby languages, typically based on individual decisions of their speakers, slowly give way to other languages. In the article you mention, I related the current process to the Neolithic Revolution: 'Evidently, what we are witnessing is not just a massive die-off of the world's languages, it is the final act of the Neolithic Revolution, with the urban

agriculturalists moving on to a different, digital plane of existence, leaving the hunter-gatherers and nomad pastoralists behind.’

What are the ideal conditions under which languages acquire digital ascent? Is there a successful model of crossing the digital divide? Do efforts to revive Basque and Catalan language in Europe provide an alternative model?

Whatever conditions favour general education, from elementary up to postgraduate level are required – ‘ideal’ is hard to define, but this has to do with the willingness of the population to better their lot. Again, this is a revolution – you must stand up and be counted! The only way to stand up is by education, native language instruction gradually extending from literacy to all skills (e.g. medical, law, engineering) that require an advanced degree. For this, you need teachers (not just elementary and high school teachers, but also college professors offering courses in their native language) first and foremost. Note that the Basque and Catalan efforts are largely driven by computational linguists who already have PhD level education.

Africa and Asia are the major hotspots of language diversity. But the communities speaking diverse languages are not only digitally disenfranchised with poor connectivity, but many languages exist in only the oral form. Under these circumstances, what are the possibilities and policies required for digital inclusion?

You answered the question: since languages that only exist in oral form have no digital future, the first step is to develop literacy, beginning with a standardized orthography. Only policies that help this will have a chance for success, and even these must fight an uphill battle.

Can machine translation tools developed by Google and the National Geographic’s Enduring Voices project equip capacities of smaller endangered languages to survive and revitalize in this gloomy scenario of linguicides?

No, I don’t think so. Every little bit helps, but it is simply not possible to revitalize top down; the effort must be bottom up, starting with the speakers of the language, and again, involving intellectuals, teachers, linguists, poets, etc. etc., who take their language seriously.

You have worked on the digital re/vitalization of languages in the Indian subcontinent; is it possi-

ble to revitalize a language having a meager presence in the digital realm? Do we have a working model/s showing such possibilities? Do you see any opportunities for funding such revitalization of languages by the government or corporate entities in India or South Asia?

Let me stay silent on the funding opportunities as I know nothing of this. But a meager presence can mushroom into a far larger presence, as long as people are committed. Perhaps the best form of funding would be for equipping teachers with laptops which can record video and audio, and more important, can enable participation in digitally mediated conversations (Zoom, Skype, Teams, Google Meet, and others) taking place in the language and recorded (with the permission of all concerned of course).

Languages have evolved in different ecozones that depict the worldview of their surroundings; often, it is closely linked to livelihood. With the increasing acceptance of a digital form of communication, can languages be de-linked from their roots and flourish in the digital realm?

Yes, absolutely. This is especially meaningful for languages that have significant scholarship going back to centuries, as is the case with many languages of India. But if the livelihood is dominantly agricultural, with no written material, not even contracts, legal documents, private letters, sacred texts, etc. this is going to be next to impossible.

In India, Google is investing USD 10 billion over the next ten years towards digital empowerment of the languages. Will the profit-driven corporate sector help in conserving endangered languages? Does it contradict the reality in which their policies of propagation and domination of monoculture of languages in internet are driven by the e-commerce?

I would not be dismissive of corporate efforts. To the contrary, I think this could be a huge help. India has dozens and dozens of languages (see my paper with Prof. Bhattacharyya) that have a very good chance of making the digital transition. Even borderline cases like Rajasthianic have crores of speakers. A speaker population of 10,000 can sustain quality high schools, and a lakh can sustain college-level education. Note that Basque has 7.5 lakhs, Catalan 4.1 million. As it stands, India already has very high-quality experts to revitalize the diverse languages in the Indian subcontinent.