

Bridging the Urban Rural Digital Gap by Breaking the language barrier

Dheeraj Mohandas Pai
Indian Institute of Technology, Madras

Abstract

India saw a tremendous digital boom in the past 6 years. The Internet usage have quadrupled. The smartphone usage have drastically increased. But how has this surge has affected different demographics of the countries? How accessible are the facilities to the rural population compared to the urban counterparts? Who are the major beneficiaries? Who were the least? Why are the reason behind the acute skews between the urban and the rural divide in the Internet adoption? And what can we do about it?

It turns out that there is an acute skew in the Internet adoption between the urban and the rural populations. The rural areas still contribute to the majority of the non-users. The urban population contributes as much as 90% of total online product transactions. We conduct a brief survey on the research conducted by Bain & Company, Google, KPMG and McKinsey Global Institute to understand what is preventing the Internet adoptions. Amongst the multiple reasons behind the skew in Internet adoption we discuss about the language barrier that is causing the skew. We discuss the existing solutions, also propose the possible technological solutions and the impact it can create to the nation.

India's Digital boom since 2014

India came across a tremendous digital boom since 2014. Cost of Internet crashed in 2017. In the same year the data usage quadrupled. The Internet users have grown by 90% percent. In 2017, more than 94% of small businesses accepted non-traditional methods of payment such as debit cards, credit cards, UPI, mobile wallets etc. Mobile Apps based cab rides increased from 140 million per year in 2014 to 480 million per year in 2017 with a whopping 3.4 times improvement. Digital transactions saw a 6.7 times increase from \$3.8 Billion in 2014 to 24.8 Billion in 2018.[4]

Internet adoption in Rural and Urban India

The digital boom was quickly capitalized by the urban population while the rural population still lags behind. According to the research by Bain & Company and Google only 10 percentage of users transacting a product online are from rural areas [See Figure 1. in page 3]. While 65% of non-users come from rural areas. The research also shows that the rural users do not increase their transactions with time. While the urban users increase their transactions with time. The research divides every transactions into four categories.

1. **Non - users** : Those who do not use the Internet at all.
2. **Non-transactors** : Use Internet for unpaid services like Social Media and News.
3. **Service-transactors** : Users who use Internet for service transactions such as payment of electricity bill, Mobile recharge, Cab booking etc.
4. **Product Transactors** : Users who use Internet to buy products.

Change in usage pattern along with time: Urban vs Rural users

The urban users tend to increase the Internet usage with time while the rural users stagnate. The graph [Figure 2, shown in page 3] by Bain & Company shows the usage pattern with time.

What prevents the digital boom in rural India?

Amongst all the reasons that has been mentioned by the users the following three were the key causes that prevented Internet adoption.

1. Inadequate grievance redressal mechanism.
2. Lack of trust amongst the users.
3. Lack of local language support and content.
4. Lack of touch or feel of the product.

The case of Local language content

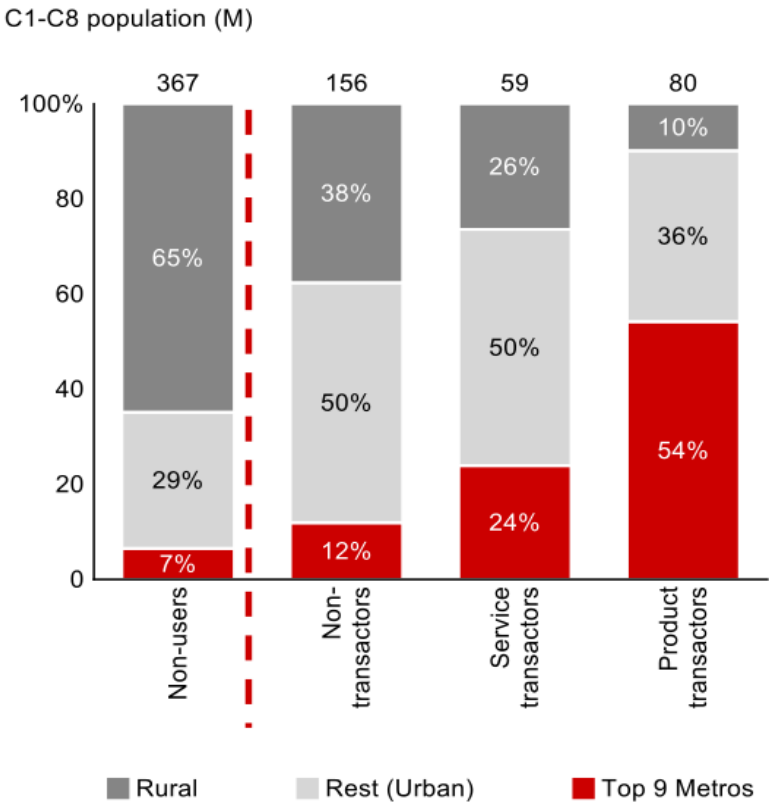
1. **Users prefer local language** : 68 % (248m) of Indian Internet prefer local language over English while transacting online.
2. **Language as a barrier** : Out of these 68% who prefer local language over 60% of them claim lack of local language content is a major barrier in product or service transaction.
3. **Limited local language websites** : While 68% users prefer local language over English there exists only <1% of total websites the user visits is in Indian local language. This is much lower compared to 3.5% in Japanese and 5% in Russian.
4. **Limited usage of local language** : 90% of the data usage in local language is consumed in chat and multimedia content.

Potential impact by bridging the language gap.

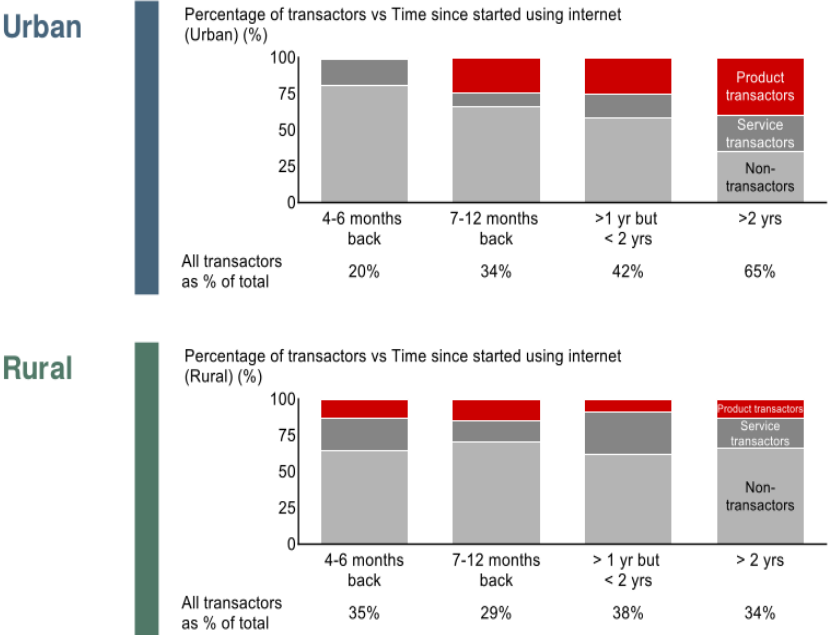
As discussed earlier more than 60% of the Indian language users feel the lack of local language support and content as the largest barrier. Bridging the language gap and improving the accessibility can significantly affects in the following ways.

1. **Increased online transactions** : The KPMG survey states that Indian language users consider content in local language as more reliable than the content in English. Around 60% consider language as the barrier. By having the content in local language could significantly boost the Internet transactions amongst the 60% of non-transactional users.
2. **Greater knowledge dispersion** : Having MOOCs, online courses, tutorials and content in local language will break the knowledge barrier that used to limit the knowledge only to the English language user.

(Figure 1 and 2)



Journey strongly reflected in urban users (especially metros), but rural users don't transition as well



Potential Solutions

Traditionally, for centuries, manual translation is used as a solution to break the language barrier. With the amount of data that is created on the Internet everyday the traditional method is infeasible and non-scalable. With the advancement in Artificial Intelligence and Natural Language Processing there are more technical solutions that are greatly scalable for the data that is created on Internet each day.

Building an AI Translation Engine contains two parts. Training and Inference. The training part is where we provide a tons of data to the algorithm and the algorithm “Learns” to translate. For Example, a translation model from English to Kannada we need terabytes of English sentences and corresponding Kannada sentences, all translated manually. Once the entire training procedure is done the model is now ready to translate other English sentence to Kannada.

Natural language translation with AI has been successfully used to translate English articles to Swedish while creating the Swedish Wikipedia. The Swedish Wikipedia is one of the greatest success of AI-based natural language translation. The Wikipedia contains more than 2 million articles making it the second largest Wikipedia.

These models can be widely used to translate tons of data in English to local language. Though the models are not always perfect. Exploitation of the structure of the language can provide additional advantage over the existing models by Google and Amazon, which are trained without exploiting any structure in the model.

References :

1. Indian Languages defining India's Internet : A study by KPMG in India and Google.
2. BCG and Google : Digital Consumer spending in India, a \$100 Bn opportunity
3. MGI Digital India Report April 2019 : McKinsey Global Institute
4. Unlocking Digital for Bharath : Bain & Company and Google

Contact details of the author :

Name : Dheeraj M Pai

Email : dheerajmpai23@gmail.com

Slides and the PDF is available at <https://dheerajmpai.github.io>