

## Localization Advantage for Indian Languages: A Brief Case Study of Tamil Trends

**S Winston Cruz, M.Phil.**

**Research Scholar, Department of Linguistics**

**Kuvempu Institute of Kannada Studies**

**University of Mysore**

**Mysuru - 570006**

**[icruz.swc@gmail.com](mailto:icruz.swc@gmail.com)**

---

---

### **Abstract**

Considerable number of Indian languages and their users are making their presence felt on the world wide web, be it content generation, content communication, ecommerce exchanges etc. A KPMG report in 2017 brought the much needed focus on the formidable presence of Indian languages in the internet space. It threw light on the fledgling Indian language user base. The report was a revelation and tacitly called for awareness and development of necessary tools and websites with multilingual capabilities in the Indian scenario. There was a natural insistence on establishing standards and implementing localization prominently that appeared as postscripts to this proposition. The scenario we have today is much bigger than what had been imagined or predicted about ten years ago, thanks to COVID-19 pandemic. Lockdowns and work/study from home restrictions catapulted the Indian languages everywhere on the internet. Online retailers and OTT platforms grabbed the opportunities. Though the trends have always been visible on various frontiers, the role played by localisation has not been elaborately studied. This study, traces some of these trends in Indian languages, particularly Tamil, that have leveraged the devices and opportunities. I will jot the trajectory by multiple markers, especially, by some of the initiatives of the software giants which may give some insights into what lies in the future for Indian languages.

**Keywords:** localization, Indian languages, Tamil, internet, ICT

### **1. Introduction**

‘We are in the middle of information era and change was disruptive and big...Future is coming sooner than anticipated’

- Prime Minister of India, quoted on *Digital India* website

The internet has become a route for achieving social and economic prospects for many nations. These countries aim at improving education, healthcare, banking, business opportunities, and communication using Information and Communication Technology (ICT). India as a nation has taken up e-governance as a potential tool for implementing its governmental schemes as well as improving the general standards of life. India launched the *Digital India* programme in 2015 with an aim to digitally empower the society and place the knowledge economy in an upward trajectory. Even before this, many private players had identified the Indian market as one their major focus areas not just for increasing their consumer base, but for expanding even their production base. The idea of ‘the biggest democracy’ caught up fast. Growing population, which was a concern for the government, was seen as an opportunity by the private companies.

The number of languages too, which has been considered as an impediment to development by some, was viewed by the private agencies as multiple paths and as new openings for reaching consumers. They considered it an efficient marketing strategy to use all the available options rather than making the public restrained to discomfort.

The private players did not have any restrictions like the Government which has different criteria like official languages, State languages, languages defined by populations or communities etc. For instance, when the Department of Electronics and Information Technology, Government of India starts developing a language tool, it has to make sure that the project covers all official languages. Any government schemes and projects through the Ministry of Education, GoI, like the National Translation Mission or the Bharatavani must consider all the 22 languages listed in the VIII Schedule of the Constitution of India first and then move to the other languages.

Whereas private players are free to approach languages according to user preferences, market perspectives etc. This is why quite a few languages, apart from the Scheduled Languages of India, find their places in the list of languages in some popular machine translators. Such a space gained by the Indian languages can be studied against different markers like localization strength of the languages, availability of standard digital tools like Unicode fonts, input and conversion software, parallel corpus and internet penetration etc.

## 1.1. Indian Languages Situation

‘Localization has become the showcase market strategy of international capitalism.’

(Pym, 2005)

Indian languages have been vying for an enviable space on the world wide web. As noted, one of the advantages for the languages is the number of their speakers. The CIA factbook (2024) lists Hindi, Bengali and Urdu from the Subcontinent among the top 10 most-spoken languages of the world. Urdu and Bengali, while being used in India as well as neighbouring Pakistan and Bangladesh respectively, use the same script in both the countries whereas Punjabi (Western) listed among the top ten most-spoken mother tongues uses the Perso-Arabic script unlike the Punjabi from India which uses the Gurmukhi script.

Six of the top 25 languages spoken as first or second language in the early 2020s were languages from the Indian subcontinent according to Ethnologue (as reported in *Encyclopedia Britannica*). According to another list, seven of the top 25 most spoken languages by native population were Subcontinent languages. The numbers are enumerated in the following table:

Table 1 - Language situation by speakers - early 2020s

Language	No. of native speakers*	Rank	No. of total speakers <sup>1</sup>	Rank	L1 speaker (%)	L2 speaker (%)
Hindi	344.65	5	609.46	3	57	43
Bengali	233.81	7	272.83	7	86	14
Marathi	83.20	13	99.22	16	84	16
Telugu	82.97	14	95.98	17	86	14
Tamil	78.59	18	86.64	19	91	9
Urdu	70.56	20	231.72	10	30	70
Gujarati	57.08	24	-	-	-	

\* All the population numbers in the table are given in millions.

While the rankings could have enthused any language enthusiast, the online presence of these languages, the focus of our study, was not very impressive at that time. One key observation for our study from this table would be the high percentage of second language speakers of

Hindi and Urdu listed in column 6. If we take here a position that L2 speakers would mostly identify themselves to be the consumers of a language and less as producers in that language, we can consider the L1 speakers as potential content creators of the language. The scenario that emerges is that the potential percentage of content producers is much higher in ratio for smaller languages in comparison with Hindi and Urdu. However, the ratio will hardly matter here as the number of both L1 and L2 population is quite high for Hindi and Urdu. Obviously, the two languages have a very high L2 population. The content here can include anything from Bollywood movies to product details appearing on an online aggregator/retailer portal to the government documents released everyday in these languages.

## **1.2. Literacy, Language Style and the Question of Content**

Literacy in India was 73% in 2011 and could have reached the 75 mark in the 75th year of its independence, that is, in 2021. Those who were aged seven and above and could read and write with understanding in any language were considered literate. The people must both read and write to be counted as literate. However, India is a country where a huge population is still living in rural areas where university level education is yet to penetrate to satisfactory levels. Only 63% of the rural adult population were literate as against 83% of the urban adult population (Ministry of Statistics and Programme Implementation | Government of India, 2016). For 1000 persons only 18 females from rural India completed their graduation as against 96 from the urban regions. For the same 1000 persons, among the males, 37 from rural areas and 126 from the urban areas graduated successfully.

Given a combination of all these statistics, one can assume that not many of the L2 speakers would be able to handle the written material in both L1 and L2 fluently. Such learnings, does not matter if they were statistically driven or by popular opinion, seems to have influenced the type of content creation for Indian language users. This is also supported by global trends where audio-visual content gains more traction than textual content (Boston Consulting Group, 2018). Entertainment content is turning almost exclusively audio-visual and so are news and information. Over the top (OTT) platforms have grown equally to the television industry offering content on various subjects from entertainment, infotainment to religious sermons and activities. Product reviews use audio visual medium and so do the culinary content developers.

In education, audio-visual medium in the form of Massive Open Online Courses (MOOCs) is popular both as informal and formal modes of education. The Government of India has its own MOOCs in the form of the SWAYAM, NPTEL and e-Pathshala portals which are slowly but steadily augmenting content to the classroom teaching. While text does accompany these videos and examinations have to strictly follow the textual path, audio-visual content generation and publishing remains the core to the users. In certain platforms, some users post in different languages and their presence on the site is a combination of their posts in more than one language platform. For example, Balaji Viswanathan on Quora (<https://www.quora.com/profile/Dr-Balaji-Viswanathan>), the question answer social media portal, is the highest subscribed globally on the portal with over five million followers. He posts regularly in English as well as Tamil.

### 1.3. Diglossia and Content

The relationship of literacy with different content types becomes interesting in the case of Tamil as it is a diglossic language. Among the two varieties of Tamil, i.e., the informal and the formal, the former is used in everyday conversation. The latter, perhaps enhanced by formal education, is mostly textual and is also followed in formal oration or reporting in television/radio news broadcasts. While the two varieties show differences at all linguistic levels in Tamil, it may be noted that there are no two separate sets of lexical items in Tamil as exhibited by many other diglossic languages. The formal variety in Tamil includes all the words which find their space in the informal Tamil but has a few more which may occur only in the formal variety. Ramaswami (1997) points out that the formal variety has two styles. The pedantic style among the two is more formal than the popular style. The popular style uses the lexical items of informal Tamil but applies the phonology of the formal Tamil. Localization efforts generally prefer the popular style in print texts or online publications. From a cursory study done for this article on the Tamil version of the ‘Google Terms of Service – Privacy & Terms’ page (Google, 2024), this researcher found a mixture of lexical pairs from both the formal and the informal Tamil used in different places of the same article, for example, *agaRRu/niikku* ‘remove/delete’, *udaaranam/eDuttukkaaTTu* ‘example’, *kuuRu/sol* ‘say/state’, *aLi/koDu* ‘provide’, *evvaaRu/eppaDi* ‘how’ etc.

The YouTubers and social media influencers producing visual content, however, prefer the spoken variety. The choice of spoken variety also erases any inhibitions about code mixing or

using words having non-Tamil roots in their presentations. They choose a style which is reasonably without regional colours, unless otherwise their channel objectives or the content demands sticking to a particular regional variety. This style is viewed as suiting their content which is neither ‘official’ nor could be termed ‘objective’.

## **2. The KPMG Report**

One can say that the report by KPMG in India (2017) from the study done in collaboration with Google about the presence of Indian languages on the internet put the Indian language user on the map. The executive summary of the report started with the lines startling everyone that Indian language users have overtaken the number of English users in 2016. The number was growing fast at a rate of 18 percent CAGR (compound annual growth rate) and was predicted to reach 536 million by 2021. This startled the pundits as well as the common public together. The Government of India had already aimed at these prospects with the *Digital India* premise. All these attracted the attention of the multinationals to increase their footprint in Indian languages. Home delivery of daily essentials during the COVID-19 pandemic acted as a catalyst. Online retailers, transport and food delivery aggregators took advantage of the curbs on public movements. Many took to online banking, online classes etc willingly while others accepted the same for lack of options. All these increased the demand for making content and apps available in Indian languages and tech giants followed it up with their own projects.

## **3. Indian Language Initiatives by the Global and Local Technology Companies**

The rise of ICT has been phenomenal in the last decade both in terms of content as well technology. A few of the initiatives began in the initial years of the last decade itself. Therefore, their success was not through any short cuts but through long thought out plans extending to the future. However, with the advent of artificial intelligence (AI) technology, especially, large language models (LLMs) and conversational generative AI (CGAI), many new players have entered the arena. Here we look at some long drawn projects to mark the trends in Indian languages on the internet.

### **3.1. Next Billion Users (NBU) from Google**

Next Billion Users (NBU) initiative was launched by Google LLC in 2015 focusing on developing a better and easier internet for users who started accessing the internet for the first time. They were using their smartphones mostly to access the internet and using the

---

**Language in India** [www.languageinindia.com](http://www.languageinindia.com) ISSN 1930-2940 Vol. 25:4 April 2025

S Winston Cruz, M.Phil.

Localization Advantage for Indian Languages: A Brief Case Study of Tamil Trends 69

smartphone itself was a new experience for them (Ranjan, 2022). New banking apps, map apps etc. were created facilitating their various daily needs. The focus on India provided the language challenge. Interchangeable user language interfaces, multimodal accessibility through video and voice for searching content etc. took the center stage of software development (Anandan, 2018). These internet users were expected to be from lower income groups, with varying standards of education and even their connectivity to the internet was relatively unstable (O'Reilly, 2005). Google clubbed this NBU initiative with its localisation efforts. Many apps under its inventory like Google Play, Google Maps etc and a co-product like YouTube were localised greatly to attract the new users and facilitate their experience on the internet.

### **3.2. YouTube Scenario - 2025**

The online video sharing platform YouTube.com is the second most visited website in the world and even in India. It was started in 2005 and Google LLC bought it in 2006. Unlike Google search engine which pulls pages from different websites and information from various databases based on the search queries from the users, YouTube search pulls videos from channels hosted on its own platform. About 15 channels in the top 50 YouTube channels were from India in 2023. In fact, an Indian music channel T-Series topped the list of YouTube subscribers in that year according to Kaggle.com statistics. While 14 of these channels were Hindi channels, Wave Music, ranked 33, was a Bhojpuri music channel. Apart from the music and film channels, one of these primarily posts children's educational content, another offered religious music and one more was a news and politics channel. The language of the channel is primarily identified as the language of the content here and not by the option of changing the interface in the particular language. However, opening up for the Indian market was long planned by the company. In 2016, Marc Lefkowitz, the then head of YouTube Content Creator Asia Pacific in an interview to an Indian business newspaper said that the company looked forward to increasing content in Indian languages by 40 percent by the next year (Chaturvedi, 2018). Many strategies were formed towards this. YouTube Music, for instance, created playlists in Indian languages according to various genres, actors, musicians etc of these languages. At present each Indian language has different scenarios in terms of top channels, content and the breed of YouTubers. It has given rise to new genres like stand-up comedy. Comedy is generally a top genre in social media and YouTube in particular. However, creating a space for artists who are not part of the film or TV industry which are the biggest sources of

comedy artists and let them become celebrities with their own standing is an achievement. YouTube influencers and motivational speakers are another tribe and product of the social media giant. In a language like Tamil, country life has got a new lease of life through channels about village cooking, natural farming etc. They are some of the most subscribed channels in the language.

### 3.3. The Wikipedia Scenario - 2020

Wikipedia, the voluntary crowdsourcing encyclopedia, is available in many languages. User participation in Wikipedia for creating and moderating the pages on different topics was a trendsetter which has been followed by many others later. Wikimedia Foundation runs this online encyclopedia and makes the content available under Creative Commons license. The Foundation complements Wikipedia with associative portals like Wikimedia, Wikimapia, Wiktionary etc. There is also a chat page to every page created where contributors converse about their choices, veracity of the information etc. The Foundation also organises workshops to engage language communities and potential contributors are trained through them. These efforts lead to better content generation. A Wiki Meta page provides information on the number of Wikipedia pages according to languages. Following is the ranking extracted by the end of 2022 which has been compared with the 2025 scenario:

Table 2 - Subcontinent Languages with 10,000+ Wikipedia Articles in 2022 and 2025

Language	No. of articles (2022)	Rank (2022)	No. of articles (2025)	Rank (2025)
English	6,562,386	1	6,971,551	1
Cebuano	6,125,564	2	6,116,768	2
Urdu	178,690	55	219,339	54
Hindi	153,338	59	165,278	63
Tamil	149,326	61	172,805	60
Bengali/Bangla	128,845	65	166,114	62
Marathi	87,483	74	99,626	74
Malayalam	79,564	78	86,685	82
Telugu	78,934	80	110,507	71
Newari	77,352	83	72,501	89

Western Punjabi	66,324	89	74,001	88
Punjabi	38,852	106	57,917	99
Nepali	32,130	109	29,907	117
Gujarati	29,997	111	30,534	116
Kannada	28,831	113	33,588	114
Bishnupriya Manipuri	25,085	117	25,087	125
Oriya/Odia	16,074	131	19,363	134
Sindhi	15,276	138	18,958	136
Maithili	13,720	145	14,192	150
Sanskrit	11,796	153	12,275	161
Assamese	10,777	158	17,660	138
Meitei	10,133	164	10,435	173
Sinhala	-	-	22,565	128
Santali	-	-	13,114	156

All the Subcontinent or Indian languages are ranked below 50 in the list both in 2022 and in 2025. Urdu and Hindi whose L1 and L2 speaker populations numbered 246 million and 609 million respectively were ranked 55 and 59 in 2022. While Urdu climbed to 54th spot in 2025, Hindi slipped to 63. Both of them had a little more than 1,50,000 articles as against English which had over 6,00,000 articles. In fact, no Asian language figures in the top 10 list except Cebuano, a language from the Philippines. The unusually high number of articles in Cebuano is attributed to the use of a bot and the articles were not naturally created by the users. It is interesting to note the place of Tamil in the list, which ranked 17 according to population and quite less in actual number of speakers compared to Hindi, yet took just two places below Hindi and four places below Urdu in the ranking for the number of Wikipedia articles. This achievement of Tamil is certainly remarkable, and it has to be studied to see if this can be replicated for other languages.

Comparing the Wiki Meta data about Wikipedia users of 2022 and 2025 one can also notice an incongruous factor that the number of active users has gone down from 311, 1389 and 349 in 2022 to 258, 973 and 315 in 2025 for Urdu, Hindi and Tamil respectively. On the surface, this seems to be in direct contrast with the KPMG (2017) report which predicted that more

Indian language users would become active users when the internet penetration increases. However, the criteria of active users is different in both cases. For Wikipedia, an active user is someone who has registered herself with Wikipedia and she should get involved in editing the Wiki pages at least once in 30 days.

### **3.4. Project Tiger**

Wikimedia Foundation is proactive in creating the contributor base to its regional language versions. They conducted a writing contest in Indian languages called *Tiger Project*. This particular project was done in collaboration with Google. It was launched with an intent to fill areas where content was less in Wikipedia. (Wikimedia, 2018). The duration of the contest was three months. The results showed that most number of pages were created in Punjabi. 32 users created 1,320 pages followed by Tamil where 56 users participated and created 1,241 pages. However, the Tamil pages had 63,617 new page views against 10,926 of Punjabi. In fact, Bengali and Malayalam which had a mere 379 and 251 articles written respectively, far exceeded Punjabi under new page views criterion with 29,740 and 18,763 respectively. Hindi had 17 users creating 280 articles which received 11,819 new page views. It is interesting to note that the contest also included English where 7 users created 20 pages which had an overwhelming response of 85,296 new page views. The project gives us an insight into how the number of speakers, availability of tools etc may influence different internet language communities differently.

## **4. The Influence of the Social Media**

WhatsApp, Facebook, Twitter, Instagram etc have certainly become inevitable sources of both entertainment and information in many Indian languages. They have helped many individuals to connect with their schoolmates, many community organisations to connect with their members, many companies to compel their workers to be available online even after the office hours etc. Students remain connected with their teachers even after school hours. The social media is credited with influencing the mass movement in 2017 for restoring the legal rights for conducting *jallikattu*, the South Indian equivalent of the Spanish matador, in Tamil Nadu (Sivaraman, 2024). As mentioned before, some of the little Davids, the individual YouTubers, have beaten many tech Goliaths already. The success of social media is attributed to cheap data charges, availability of technology in Indian languages etc.

## 5. Script and Tools

Computing tools and software arrived relatively late to Indian languages when compared with the European situation. The most used script on the internet is obviously Latin. Major languages of the world and the internet like English, Spanish etc use the Latin script. Hindi and a few other Indian languages like Bodo, Dogri, Maithili, Marathi, Nepali, Sanskrit, Santali and Sindhi use the Devanagari Script. The Government of India also has a policy to introduce writing in Devanagari in any language that has not been reduced to writing yet. The Sahitya Akademi, the National Body of Letters of India, recognises only Devanagari for a language like Konkani which is written in multiple scripts in different States of India. Each of the major Dravidian languages use their own scripts and occupy different code blocks and so do Assamese, Bengali, Gujarati, Odia, Punjabi etc. Some State Governments had their own departments developing fonts and input systems and tools for the State languages from pre-Unicode times. A few non-governmental organisations and many individuals too have continuously contributed towards software development in Indian languages.

The Unicode Consortium adopted the ISCII system developed by the Government of India (Sengupta & Mooney, 2019) for Indian languages. Microsoft in their Windows OS included the Inscript keyboard initially for Indian languages along with the phonetic keyboard layout for some. Later it also included other standard keyboard layouts available for some languages, e.g., Tamil99 keyboard layout developed by the Government of Tamil Nadu for Tamil. Rajendran (2006) provides an annotated list of most of the tools that were available till 2006 for Tamil, just around the time Unicode became popular for Indian languages. Such tools have come in handy in the use of these languages on the internet.

Of late, Gboard keyboards are available for Indian languages on Android smartphones and they also allow voice typing. The latter has become a real game changer in Indian languages. Many people use the voice typing facility even though they know the keyboard layouts and are comfortable with typing in these languages. Apart from typing, image to text and translation tools etc also allow the use of Indian languages in a big way.

On the other hand, deep learning techniques in machine learning have made it possible to create translation systems and other language tools even with smaller corpus. Conversational Generative AI has not left Indian languages far behind. Llama2 is the latest of the Large

---

**Language in India** [www.languageinindia.com](http://www.languageinindia.com) ISSN 1930-2940 Vol. 25:4 April 2025

S Winston Cruz, M.Phil.

Localization Advantage for Indian Languages: A Brief Case Study of Tamil Trends 74

Language Models released by Meta AI, the owners of Facebook, WhatsApp and Instagram. It has become a source for many open source AI initiatives and individual researches on Indian languages. Meta itself has embarked on a mega initiative called 'No Language Left Behind' (NLLB) for all world languages and the greatest advantage of it is that it is using the open source models (Costa-Jussà et al, 2022). This will be the trend for the future. All developments in language technology need to be inclusive and Indian languages are to be covered by most of them.

## **6. The Trends and Conclusion**

It is now a fact that Indian languages thrive on both the internet as well as the artificial intelligence technology scenario. Large Language Models (LLMs) are being made available with Indian languages. Popular AI systems include Indian language translation, information retrieval etc. Both the public and private sectors are involved in these endeavours. It must be noted that though content creation in Indian languages aids these processes, the two are not well balanced. Many Indian languages lack full-fledged content creators. Indian language content creation has not become a full-scale remunerative profession except for some journalists and popular YouTubers.

At the same time, the number of content consumers are on the rise. This has led to publishing of translated content from English and other languages which are translated through either human or machine translators. The machine translated content mostly appears as on-the-fly translations. Some websites like [www.amazon.in](http://www.amazon.in) provide a user interface with human translated menus, buttons etc. However, their product details are mostly on-the-fly translations from their English content as they form a huge chunk considering the number of products that are available on their website. This leads to a confused state as far as user experience is concerned as these translations create many hallucinating moments. On the fly machine translations are also available on many service websites and other websites of the Government like [www.cowin.gov.in](http://www.cowin.gov.in), [www.uidai.gov.in](http://www.uidai.gov.in) etc. While the space for Indian languages is increasing everyday on the internet, many do not seem to capitalise on the advantage yet.

Only big players like Google, Amazon, etc. and some individual YouTubers have reaped the best out of these available tools and consumer interest.

Some Indian languages with larger populations have not been able to convert that into increasing their presence on the internet. Content creation, therefore, seems to have been taken up in some languages, for instance Tamil, not because of the size of its population. It is not an effort to establish digital capitalism. It has been taken up because the language users seem to be language lovers in the first place. It is their love to see their language everywhere that seems to fuel the high rate of development of these languages on the internet. It is also aided by the available tools as seen in the Project Tiger case where more new page views were witnessed in the case of Tamil with comparatively fewer number of pages and users. One can only agree with A Ramasamy Gounder's statement in the December 1943 issue of *Tamil Pozhil* magazine (reported in Venkatachalapathy, 2000) who describes the phenomenon not as any hate towards any other language but as the love and passion for one's own language. This will be the crucial element for Indian languages to stay relevant in the digital future.

=====

### References

- Anandan, R. (2018). *Google for India: building services for every Indian, in their language*.  
Google. <https://blog.google/technology/next-billion-users/google-for-india-2018/>
- Boston Consulting Group. (2018). *Entertainment goes online*. Retrieved March 8, 2025, from  
[http://image-src.bcg.com/Images/Entertainment-Goes-Online\\_tcm21-208006.pdf](http://image-src.bcg.com/Images/Entertainment-Goes-Online_tcm21-208006.pdf)
- Central Intelligence Agency. (2025). *The world factbook 2024*. CIA. Retrieved April 1, 2025,  
from <https://www.cia.gov/the-world-factbook/countries/world/>
- Chaturvedi, A. (2018). *YouTube will speak in your language more*. The Economic Times.  
Retrieved March 4, 2025, from <https://economictimes.indiatimes.com/industry/media/entertainment/youtube-will-speak-in-your-language-more/articleshow/66079094.cms>.
- Costa-Jussà, M. R., Cross, J., Çelebi, O., Elbayad, M., Heafield, K., Heffernan, K., ... &  
NLLB Team. (2022). No language left behind: Scaling human-centered machine  
translation. *arXiv preprint arXiv:2207.04672*.

- Encyclopaedia Britannica. (n.d.). *Languages by total number of speakers*. Encyclopaedia Britannica. Retrieved [24.03.2025], from <https://www.britannica.com/topic/languages-by-total-number-of-speakers-2228881>
- Google. (2024, May 22). *Google sevai vitimuraikaL – taniyuRimaiyum vitimuraikaLum – Google*. Google Policies. Retrieved September 10, 2024, from <https://policies.google.com/terms?hl=ta&fg=1>
- Government of India. (2011). *Census of India 2011, language - India, States and Union Territories*. Retrieved September 06, 2024, from [https://censusindia.gov.in/nada/index.php/catalog/42458/download/46089/C-16\\_25062018.pdf](https://censusindia.gov.in/nada/index.php/catalog/42458/download/46089/C-16_25062018.pdf)
- Kaggle. (n.d.). *Global YouTube statistics 2023*. Kaggle. Retrieved March 4, 2025, from <https://www.kaggle.com/datasets/nelgiriyeewithana/global-youtube-statistics-2023>
- Meta contributors. (n.d.). *List of Wikipedias*. Meta-Wiki. Retrieved March 15, 2025, from [https://meta.wikimedia.org/wiki/List\\_of\\_Wikipedias](https://meta.wikimedia.org/wiki/List_of_Wikipedias)
- Meta contributors. (n.d.). *List of Wikipedias*. Meta-Wiki. Retrieved November 30, 2024, from [https://meta.wikimedia.org/wiki/List\\_of\\_Wikipedias](https://meta.wikimedia.org/wiki/List_of_Wikipedias)
- Ministry of Statistics and Program Implementation | Government of India. (2016). *Women and men in India-2016 - literacy and education*. Retrieved September 06, 2024, from [https://mospi.gov.in/sites/default/files/reports\\_and\\_publication/statistical\\_publication/social\\_statistics/WM16Chapter3.pdf](https://mospi.gov.in/sites/default/files/reports_and_publication/statistical_publication/social_statistics/WM16Chapter3.pdf)
- O'Reilly, T. (2005, September 30). *What is web 2.0*. O'Reilly. Retrieved April 5, 2025, from <http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>
- Pace, Jonathan. (2018). The concept of digital capitalism. *Communication Theory*. 28. 10.1093/ct/qt009.

- Pym, A. (2005). Localization: on its nature, virtues and dangers. *SYNAPS 17*. NHH. 17-25.  
Retrieved August 20, 2021, from <https://openaccess.nhh.no/nhh-xmlui/handle/11250/2394909>
- Rajendran, S. (2006). A survey of the state of the art in Tamil language technology. *Language in India*, 6(10), 1-43.  
<http://www.languageinindia.com/oct2006/languagetechologyintamil.html>
- Ramaswami, N. (1997). *Diglossia: Formal and Informal Tamil*. Central Institute of Indian Languages.
- Ranjan, P. (2022, March 15). *An anthology of insights, for a more inclusive internet*. Google Blog. Retrieved April 5, 2025, from <https://blog.google/technology/next-billion-users/anthology-insights-more-inclusive-internet/>
- Sengupta, G., & Mooney, J. (2019, April 11). *Indic scripts and the internet*. Dibyajyoti Ghosh / by researchers@work / r@w blog. Medium. Retrieved April 5, 2025, from <https://medium.com/rawblog/indic-scripts-and-the-internet-739453610992>
- Sivaraman, R. (2024, August 22). The seven-day jallikattu protest on the Marina. *The Hindu*. <https://www.thehindu.com/news/cities/chennai/the-seven-day-protest-on-the-marina/article68545544.ece>
- The Hindu Businessline. (2018, December 10). *44 Tamil channels crossed 1M subscribers on Youtube*. The Hindu Business Line. Retrieved April 5, 2025, from <https://www.thehindubusinessline.com/news/variety/44-tamil-channels-crossed-1m-subscribers-on-youtube/article25711857.ece>
- Venkatachalapathy, A. R. (2000). *Antha kalathil kappi illai muthanala aaivu katturaigal*. Kalachuvadu Publications.
- Wikimedia. (2018, 02 15). *Project Tiger editathon 2018*. Wikimedia Meta-Wiki. [https://meta.wikimedia.org/wiki/Project\\_Tiger\\_Editathon\\_2018](https://meta.wikimedia.org/wiki/Project_Tiger_Editathon_2018)