REVERIE
Language Technologies

DIGITAL INDIAN LANGUAGE REPORT

ISSUE #2

H1, 2018

Contents

Democratizing the Indian Internet

500 million
Indian internet users are projected to be online by June 2018

90 %
of new internet users will be Indian language speakers

73 %
of Indian internet users will be non-English users by 2021

Trading multiple identities for one monolithic identity, that of English, is something that never made sense to us, especially since English only represents a small, privileged segment of the country's teeming masses. If India's strength is in its diversity, that diversity needs to be represented in the digital world too, for our people to be able to call the digital world their own.

By promoting linguistic quality on the internet, we're democratizing the internet.

Source:
Indian languages - Defining India's Internet, KPMG

India is home to one of the most diverse populations on Earth. The sheer number of different cultures and languages that co-exist, side by side, is what makes the fabric of our society unique.

Indians speak a wide range of languages, and only 10% of Indians speak English. This heterogeneity is what makes it a difficult market to operate in. Companies, especially global players, find it difficult to scale in this market and capture its full potential.

The Indian internet is also an expression of how diverse the country is. Around 500 million Indians are projected to be online by June 2018. Most of these users, over 73% of them, will use the internet in their own language. This number will grow to over 530 million Indian language internet users by 2020.

This internet revolution has been made possible thanks to greater internet penetration caused by plummeting data prices and cheaper mobile handsets.

While the numbers clearly show a significant shift in usage habits, companies are still yet to wake up to this new reality. Most apps and services on the Indian internet are not localized, fonts and rendering solutions are generally patchy and inaccurate, and interfaces are not designed with the Indian language user in mind. 70% of Indians have issues with using English keyboards.

99% of Indian language users access the internet from their phone, which is why we've chosen to highlight the consumer insights of mobile internet users in our report.

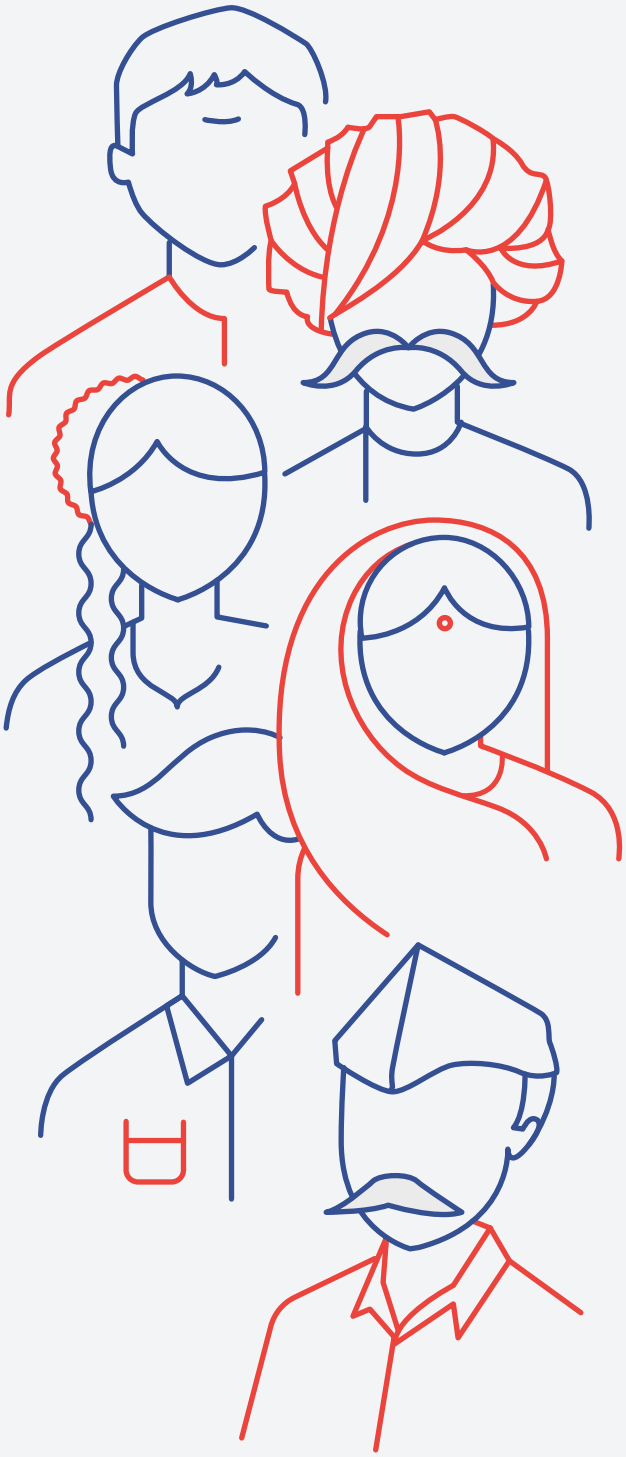
At **Reverie**, our work revolves around ***building Indian language solutions,*** ensuring that their speakers aren't left behind and at a disadvantage when it comes to using the internet and accessing essential services and information. We have witnessed that organizations that prioritize Indian language localization and make it an integral part of their India growth strategy, see great success.

We spend a lot of time trying to understand the people our work affects better, so that we can serve them and solve the issues they face better. Companies building solutions for India can use these insights to tap into the Indian language user base through language tech.

This, the ***second edition*** of our Digital Indian Language Report, represents yet another step in our never ending journey to learn more about how the average Indian interacts with the digital world and the role of language in their equation. It represents a record of what we've learned so far.

Arvind Pani,
Co-founder & CEO,
Reverie Language Technologies

Consumer **Insights**



Top Indian Languages Online

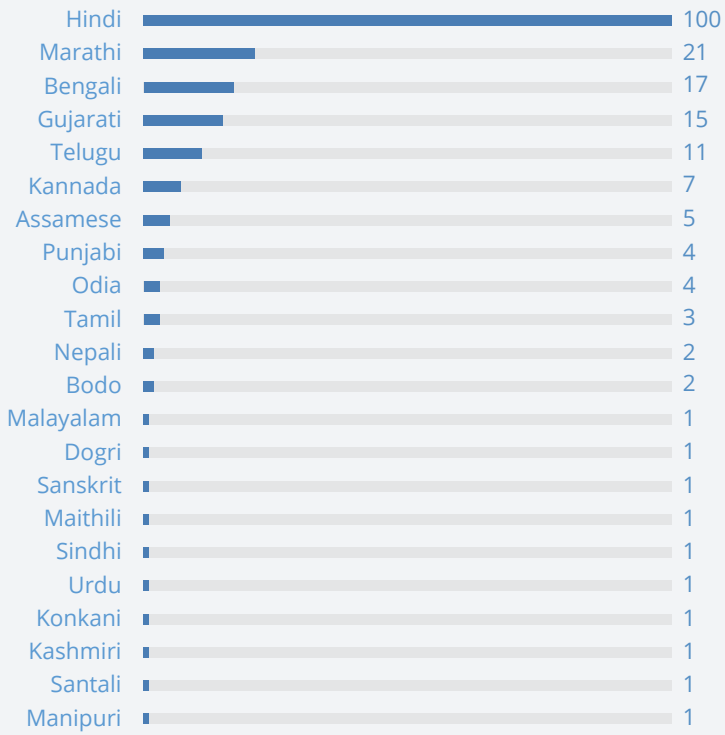
Out of the 22 languages studied, ***Hindi, Marathi, Bengali, Gujarati*** and ***Telugu*** saw the most use online, by words typed.

The ***Hindi user base*** is comparable to that of the rest of the languages user base combined, indicating a much higher number of words typed. ***Telugu*** and ***Kannada*** are the most used Dravidian languages.

Linguistically, individual words in Dravidian languages and Marathi can correspond to entire phrases with multiple words in other languages. Despite this fact, these languages have a high number of words generated.

Total number of words generated by language

(Indexed to 100 for highest)



Engagement Levels

The engagement level of an Indian language corresponds to the ***average number of words typed per user*** in that language. This measure differs from the measure of top Indian languages online in that it gauges the average user's engagement with the digital medium.

Marathi, Bengali, Telugu, Gujarati and ***Hindi*** users show the highest levels of engagement, in that order. This data point shows us that users of these languages, especially ***Marathi*** and ***Telugu*** which already have a inherent linguistic tendency towards longer words, show a greater propensity to engage with digital services and platforms, driving usage beyond mere passive content consumption, into active usage territory.

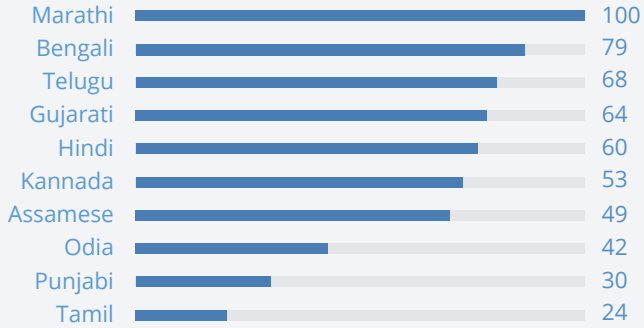
Minority Languages

Apart from India's major languages, research shows us that minority languages are also widely used online. The internet allows speakers of minority languages to use their native languages in casual writing, without needing them to use specialized language like in literature or journalism.

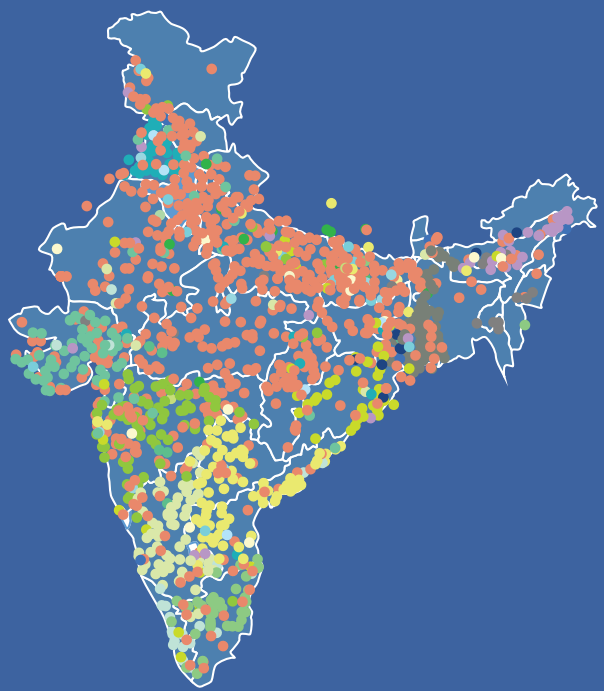
Bodo, Dogri, Maithili, Sindhi and ***Santali*** are the most widely used minority languages online, in that order.

Average number of words generated per user

(Indexed to 100 for highest)



Data clearly shows us that there is a very real preference for speakers of these languages to communicate in their respective mother tongues online. Linguistically speaking, this also bodes well for the languages as a whole, *as the digital medium facilitates the production of content and ensures their continued survival.*



Language And Region

Indians have been migrating across the length and breadth of the country for centuries now. In modern times, Indians have been moving for work more than ever before. This migration has also given rise to the wide dispersal of languages across the country, and this is reflected by our data. Linguistic islands of digital use sprout up all over the map, representing Indians who have either moved to different linguistic regions, or linguistic minorities using their own languages.

Assamese	Konkani	Sanskrit	Bengali	Maithili
Bodo	Malayalam	Santali	Sindhi	Dogri
Gujarati	Manipuri	Tamil	Telugu	Marathi
Hindi	Nepali	Urdu	Kannada	Odia
Kashmiri	Punjabi			

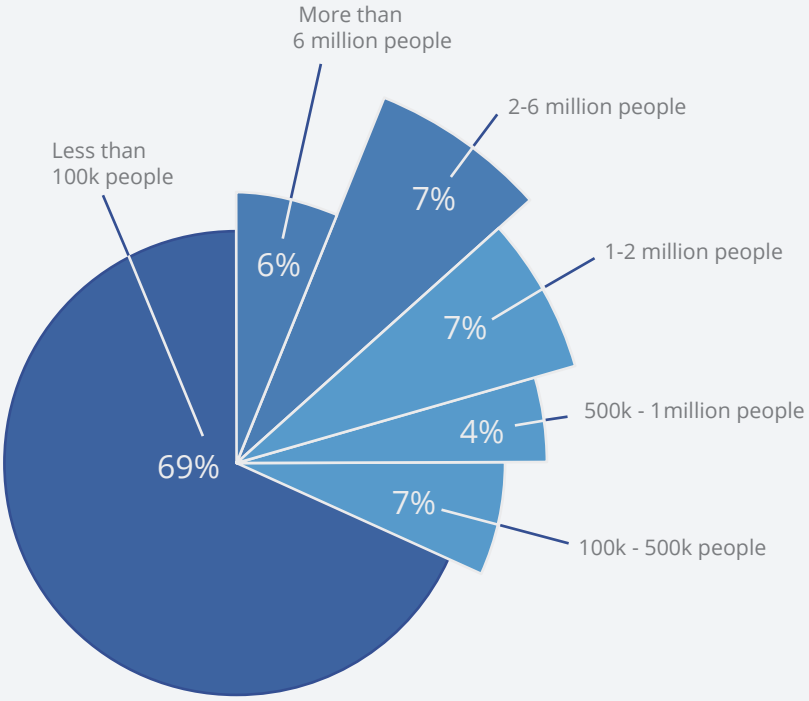
This data shows us that a one region one language formula for localization, usually using geo targeting, might not be an ideal strategy.

Demographic Breakup
- Population

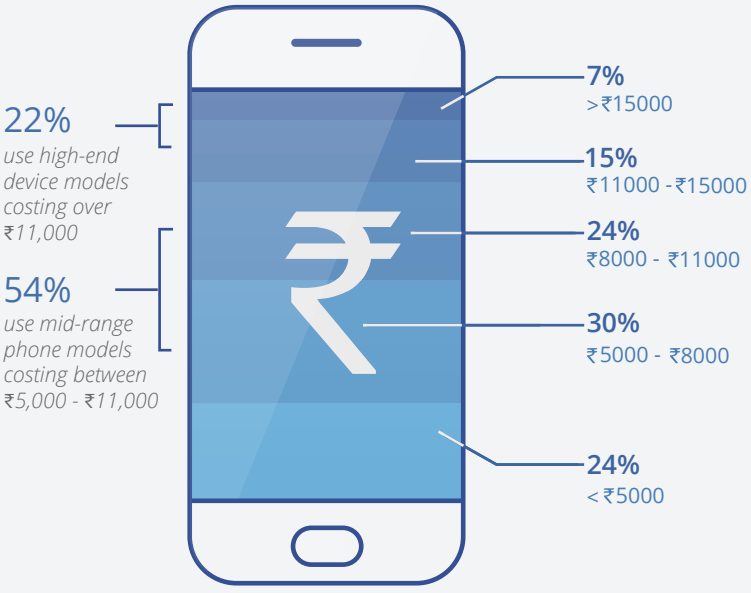
The majority of Indian language internet users at **69%**, are from towns and villages with fewer than 100,000 people. **13%** of users are from larger cities with more than 2 million inhabitants. Users from smaller cities and medium sized towns with 100,000 - 2 million inhabitants make up the remainder, at **18%**.

This ***large city vs small town + rural split*** has significance when it comes to localization perspectives as well. Lifestyle differences exist between the two classes, meaning that *in order for localization to truly be effective, it must account for these differences in addition to the Indian languages used.*

For example, users from small towns may not understand what a shopping cart icon denotes. They would be more familiar with a shopping basket instead.



Breakup of Indian Language
Internet User base by Town Class



Breakup of Indian Language
Internet User base by Device Price Band

Demographic Breakup
- Device Price Band

The majority of Indian language internet users, at 54%, use mid range phone models.

22% of users use high end device models, proving that the assumption companies have that Indian language internet users have low purchasing power is ultimately inaccurate.

This data shows us how essential it is for phone manufacturers and OEMs to ensure that phones across all price bands have accurate, legible fonts & rendering in Indian languages for a seamless user experience.

Top App Categories

Unsurprisingly, the top apps used by Indian language internet users all correspond to low language friction categories, like social media, streaming and browsers. These users are still not entirely mature users in the way English language users tend to interact beyond content consumption and low level content generation.

This is also an indicator that higher friction categories still do not provide a smooth experience in Indian languages, leading Indian language internet users to drop off. This also applies to low language friction categories with requirements like better font & rendering support that are ignored, like gaming.

Top App Categories by User Reach

With top Indian languages and top apps used

1. Communication
Hindi, Marathi, Gujarati

2. Social
Hindi, Bengali, Gujarati

3. Tools
Hindi, Bengali, Marathi

4. Video Players & Editors
Hindi, Bengali, Marathi

5. Personalization
Hindi, Gujarati, Bengali

6. Shopping
Hindi, Bengali, Gujarati

7. Travel & Local
Hindi, Marathi, Gujarati

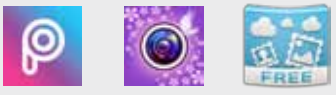
8. Productivity
Hindi, Telugu, Marathi

9. Entertainment
Hindi, Bengali, Marathi

10. Music & Audio
Hindi, Bengali, Marathi

11. Finance
Hindi, Gujarati, Marathi

12. Photography
Hindi, Gujarati, Marathi



13. Books & Reference
Hindi, Bengali, Marathi

14. Maps & Navigation
Hindi, Marathi, Bengali

15. News & Magazines
Hindi, Marathi, Bengali

Top Android apps used by Indian language users

1. WhatsApp
2. YouTube
3. Google
4. Facebook
5. Facebook Messenger
6. UC Web Browser
7. Chrome
8. Facebook Lite
9. Google Maps
10. Opera Mini
11. Gmail
12. UC Web Mini
13. Truecaller
14. Amazon
15. Instagram
16. Flipkart
17. Paytm
18. Jio Music
19. Jio4GVoice
20. Hike Messenger
21. Hotstar
22. JioTV Live
23. IMO Call & Chat
24. OLX
25. Clash of Clans

- *7 of the top used apps are messaging and social media apps.*
- *4 of the top used apps are video and music streaming apps.*
- *4 of the top used apps are shopping apps.*
- *Only one of the top used apps is a gaming app.*

Broken Language Experience

These stats raise some questions that the Indian tech industry needs to answer. It's already been established that Indian language internet users form the majority of internet users on the Indian internet. The industry, however, is yet to catch up and solve the pain points these users face.

Most Indian language internet users are still new to the internet and are not mature users yet. Only 0.1% of content online is in Indian languages, and users still face difficulties in inputting and viewing Indic text, to say nothing of actually interacting with Indic content and making transactions.

Methodology

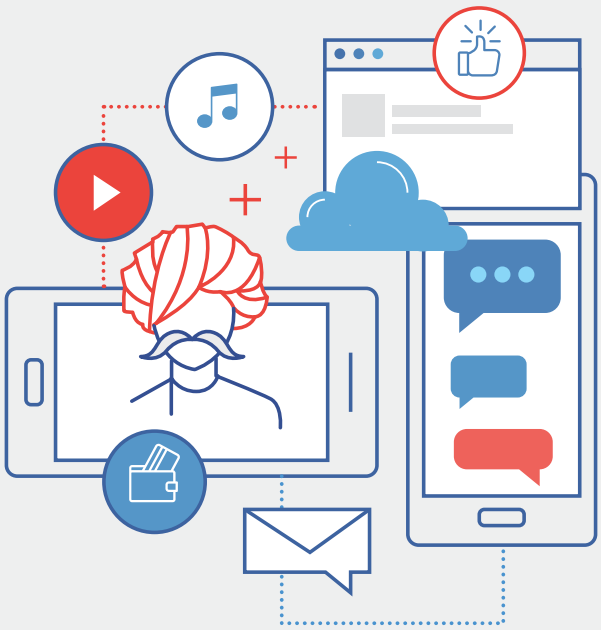
These insights have been drawn from Reverie's Indic Keyboard ***Swalekh Flip*** Android app aggregate data from July to Dec 2017. No sensitive or personally identifiable information that can be attributed to individual users has been used in this report. This study was conducted using a sample size of 89,000 Android users, who installed ***Swalekh Flip*** voluntarily from the Google Play Store. This was to ensure random sampling and that the data was representative of the Indian market. For further queries, please feel free to write to us at : **mrkt@reverieinc.com**.

Scan QR code for Indic Keyboard Swalekh Flip by Reverie

Business **Perspectives**

Image originally posted on : https://qz.com

Most Indian internet users are still not completely mature users, and their user behavior is characteristic of different stages of an internet user's user journey. This journey begins with physical acquisition of a mobile device, followed by connecting to the internet.



It's at this point that they join the ranks of the ever growing mass of hundreds of millions of Indians using the internet for everything from news, messaging, social media, gossip, music downloads, train bookings, payments, video streaming and more.

Content Consumption

Once a user has come online, they start out as *a passive consumer of content*. They read, share, watch, listen - and their level of interaction with the internet ends there. The usage patterns of a large segment of Indian internet users fall into this category, thanks to its extremely low friction nature.

The fact that so many Indians are content consumers could also tell us that *user education is still lacking*, and that *platforms can offer their users some help with that*. It could also be a sign that current localization solutions are insufficient for these users to become *content creators* rather than *passive consumers*.

Messaging and social media are the most used verticals by Indian language users

Companies are focusing on users that fall in this category, by encouraging to interact with platforms more, use services more, and create more content.

Engagement

Next comes the engagement stage, where users actively engage with platforms. Users go on to actively get more out of the platforms they use, allowing for a wide range of possibilities, restricted only by the nature of the platform itself.

Lots of low friction apps and verticals involve this stage of consumer maturity. ***Messaging and social media are the most used verticals by Indian language users*** and they correspond to this stage.

Companies are focusing on users that fall in this category, by encouraging to interact with platforms more, use services more, and create more content.

Transaction & Service

The **Indian language internet user base** is sizeable, but businesses cannot adequately target them with true local language experiences just yet.

Banking, e-commerce and other transaction related verticals, don't involve much content creation. Apart from basic interface localisation, these verticals require a number of other processes to come together to provide a user with a complete localized experience.

Until then, the Indian language internet user base cannot achieve true maturity and transact online.

India's Language Mandate For Mobile Devices

Vivekanda Pani, Co-Founder & CTO

The Indian Govt. has mandated support on mobile devices for

22 official scheduled Indian languages

This mandate will come as a relief to hundreds of millions of phone users across the country, half of whom own devices that are incompatible with Indic languages.

The Indian government, with the help of its Bureau of Indian Standards arm, has mandated support for all 22 official scheduled Indian languages on all phones sold in the country – both feature phones and smartphones. In addition, they must adhere to certain defined character sets for each language, and they must include input solutions for *at least 2 Indian languages.*

When different devices use different character sets, you end up with inconsistencies across devices which affect how the text renders on your device. Unicode is script based, which means it includes multiple characters that may or may not be used in the languages that use the script in question (***like ऍ, which does not exist in Hindi***). It also includes archaic characters and junk characters. A standardized character set does a lot to remove these inconsistencies.

By mandating uniform character sets for each language, junk characters and various other font inconsistencies are dealt with, and display and rendering issues are resolved. Somewhere along the line, being able to effortlessly use the internet and your smartphone became a privilege only afforded to English users. The rest of India has had to wait, unceremoniously excluded, for over a decade. But not for much longer.

This mandate will come as relief to hundreds of millions of phone users across the country, half of whom own devices that are incompatible with Indic languages. This massive chunk of Indians left without the ability to take advantage of what the internet or their smartphone has to offer will finally see some respite. Half the country's phone users literally have next to no clue what information they're receiving on their phone, since it's all in a language they do not know, *an elite language – English.*

With access to the internet becoming less intimidating, and more democratic, India's internet user base will see a robust growth in the coming years, a growth that will translate into ***increased demand for Indic language content*** *and* ***services available in Indic languages.***

Within a year of the mandate's release, devices that support Indic languages will form the majority of devices in circulation.

With every passing quarter, the number of phones with Indian language support will steadily climb, bringing with it an increase in Indic content and support across apps and website, creating a multiplier effect. This is a very positive development for the Indian internet as a whole, as well as for every Indian looking to access more information.

Digital Citizen Engagement

Mithun Das, *VP - Sales, Government Relations*

The internet has removed a lot of hassle by making the logistics of citizen engagement a whole lot simpler. It helps reduce the length of queues at government offices.

The Government of India has committed itself to building a **Digital India** for its citizens, with e-governance and digital government services playing a major role in this vision. For ***Digital India*** to truly reach each and every Indian and be accessible, this vision has to be built in languages the average citizen speaks.

Since English is only spoken by ***10% of India***, limiting these engagement platforms to English is to fundamentally limit their reach. Indian languages on the other hand, truly democratize any service that incorporates them.

As government agencies seek to build and establish their reach online by providing citizens with ***easy-to-use***, ***reliable digital services***, they need to design their sites keeping these issues in mind.

Since English is only spoken by

10% of India,

limiting these engagement platforms to English is to limit their reach.

A Chief Minister's dashboard where citizens can review the performance of a government, for example, shouldn't provide a Kannada language user with results in English. That defeats the whole purpose of having an interface in Kannada.

Localize Interactivity

Building platforms for citizen engagement goes beyond translating the front-end of a website or an app, however. It should cover all aspects of an interface with no gaps in language.

Think of all the possible actions a user could take to interact with a page. You have pop-ups, hover text, drop down menus, text boxes, reports, data, response forms, and even real-time updates - they all need to be localized.

Plug Leaks In User Experience

A citizen can visit a website or open an app, enter their details, select options from lists and menus, fill in text forms, select options and checkboxes, and submit these details to the server.

Ideally, this process should involve both parties, on both ends of the equation – the platform's front-end that the citizen interacts with, and the platform's back-end that processes these requests and provides output.

When a user uses a platform in a certain language, it is expected that all points of interaction will be in the user's chosen language.

Fix Content & Display Issues

Sometimes, text is incompletely localised. Several terms are left in English, or merely transliterated into the language's script. ***Consistency*** and ***accuracy*** in localised content is something that needs to be checked and examined for quality.

Display in different languages is often inconsistent. This is often the fault of inferior fonts & rendering. When text is messy and illegible, it makes the platform's accessibility poor as well, severely limiting the degree of interactivity a citizen can have with the platform.

Train Bookings - A Case Study

According to Reverie's report findings, government train booking apps (IRCTC & NTES) are among the most popular travel apps in India. However, they only reach a minority of India's 500 million strong internet user base. If they were localized in Indian languages, their reach would expand.

The Importance Of Quality Indic Fonts & Rendering for Digital Devices

SK Mohanty, Co-Founder & Chief Font Architect

The display quality of text highly influences a reader's perception of content. If a user finds it difficult to read content, they reject it no matter how good the content is. Illegible content is a common problem faced by users of languages that feature complex scripts, like Indian languages. This is why the usage of Indian languages digitally has not taken off, in comparison with English, Chinese or Japanese.

Text Display Quality

Display quality for text on digital devices hinges on the quality of the font used, and the accuracy of the rendering engine used to render these fonts. Both these elements come together to visually define the nature of the text.

India is a country of incredible diversity. Its population of 1.3 billion speaks many distinct languages, out of which 22 major languages have constitutional status. These languages are used for *day-to-day communication*, as well as in the fields of *education, literature, journalism, science, technology, e-commerce,* and more, with large volumes of written texts.

The Indian Language Display Experience

Unfortunately, we do not have optimal text display solutions for Indian languages on digital devices. Most websites with content in Indian languages are virtually unreadable and aesthetically extremely poor. Their low quality drives away users, and provides a strong incentive for them to switch to English language content instead. Most mobile phones are used for voice-based applications, despite them being more than capable of supporting text based applications.

As a result, companies & organizations cannot reach out to their target audience in their own language online, missing out on a large market opportunity, a potential user base of around 300 million people. The possible channels are vast and diverse, and include banking, e-commerce, e-governance, entertainment, news, instant messaging, and more.

If a user finds it difficult to read content, they reject it no matter how good the content is.

क्वालिटी

What Inaccurate Indic Fonts Mean for Users

Many companies develop applications for digital devices, which piggyback on the existing OS and platforms primarily designed for simple scripts like the Roman script. If the underlying display solutions cannot support complex Indic scripts faithfully, these apps will pick up those same flaws.

In India, most Indian language content has been forcefully fit into applications that are primarily designed to support simpler scripts like the Roman script. Often, Indian users find it difficult to consume Indian language content, either because such content is illegible, the text does not render properly on the device used, or because substandard fonts are used.

Because of the government's recent mandate for digital payments, transactions and various e-governance projects, there is now a rapidly growing requirement for Indian language friendly digital devices. Apart from this, there is an even bigger demand for digital communication, content consumption, digital learning, and e-commerce in Indian languages.

This will only become a reality if these devices are equipped with quality fonts, capable of providing error-free rendering, allowing their users to use their own language digitally, as effortlessly as they speak it.

*Display quality for text on digital devices hinges on the **quality of the font used**, and the **accuracy of the text rendering** used to render these fonts, depending on what suits the medium used better.*

OpenType Fonts Specially Designed for Smartphones *(in 22 Indian languages)*

Scalable Indian language fonts for smartphones. Now users can read a full newspaper in Indian languages on a tiny mobile phone screen with ease. Reverie's multilingual fonts provide unmatched user experience with **highly legible fonts** and **100%** accuracy in text rendering in complex scripts.

Fonts & Rendering Solutions for Feature Phones *(in 22 Indian languages)*

Reverie offers Indic fonts and rendering solutions on feature phones in all 22 official Indian languages. Now complex Indic scripts can have expressive text quality even on low resolution screens.

Video & Music Streaming Platform Localization

Rajesh Mehta, *VP - Enterprise Business*

300 million	90%	14%
Indian language internet users today	already use digital platforms for entertainment	of an average Indian netizen's time is spent online on entertainment

According to Reverie's own research, ***Hotstar*** *&* ***Gaana****, both Indian streaming platforms focused on Indian language content are among the top apps used by Indian language internet users. This is an example of how Indian language content boosts usage.*

Indians love their entertainment - both in video and music format. The internet has made easy, instant access to entertainment possible, through streaming. Streaming lets a user access content in real time. All they need to do is select their preferred content from a library, or search for what they want, and hit play.

Although most media consumption is in Indian languages (as evidenced by the demand for Indian language movies and music) most major streaming platforms are not localized, and are English only.

Interface Localization

The first point of contact any user has with a platform is the platform's own interface. Opening the platform should first prompt the user to select their preferred interface language. ***Language selection buttons should feature prominently.***

Streaming platforms tend to have short text strings, mostly listing genres and titles & artists, with the rest taken up by preview buttons and promotional images. Menus and pop up prompts, as well as notifications add to that.

These are mostly static text strings, and can easily be localized by translating them. However, artist names and titles are proper nouns, and need to be transliterated instead, converting only the script.

Summaries & Descriptions

Pages generally feature synopses and descriptions of content, including summaries for videos, short news snippets, explanations for features, and more. These generally follow the same rules as interface localization, but the content here differs in text volume.

Localizing can give streaming platforms the giant push they need.

This localization, however, needs to account for the nature of these streaming platforms themselves - search dependent, with lots of static content and proper nouns.

Indic Search & Content Discovery

Streaming platforms rely heavily on search and discovery. With their massive libraries, it would be impossible to display everything at once. Search helps users make sense of their options and lets them find what they want.

In order to search for content, you need input options that lets you type in the language you want to search in. A search bar should have a built in keypad for the selected language, allowing users to type directly in the language of their choice.

For search to support ***Indic typing*** and ***queries*** in Indian languages, the engine needs to index pages such that they can be discovered no matter what language or script is used.

Caption Localization

People don't only watch and listen to movies and music in their own language. They sample and enjoy them from other cultures, other languages as well. That's where captions come into the picture. They help users understand content that is not in their own language. Captions help streaming platforms because it helps them ensure that content made for one user base can be enjoyed by all users.

Unity Rendering For Indic Scripts

Rishi Kudale, Marketing Head

Unity is one of the most powerful and popular tools for game development across platforms, and is widely used by indie game developers. Originally used to make games for desktops and consoles, *it now lets you build games on 27 different platforms, including smart mobile devices.*

Mobile Gaming In India

According to a report by Flurry Analytics, India has emerged as one of the top five countries worldwide for online mobile gaming. Indian gamers are responsible for 13% of all global gaming sessions, second only to the US's 20%. Mobile gaming revenue in India is also expected to reach $1.1 billion by 2020.

While these stats are impressive, they are mostly restricted to an English user base – India is home to nearly 1.3 billion people, and only around 10% of them know English. In addition, more than 50% of India is below the age of 25, and more than 65% of Indians are below the age of 35 - the prime age groups for gaming. India is home to the second largest number of developers in the world.

Gaming Localization

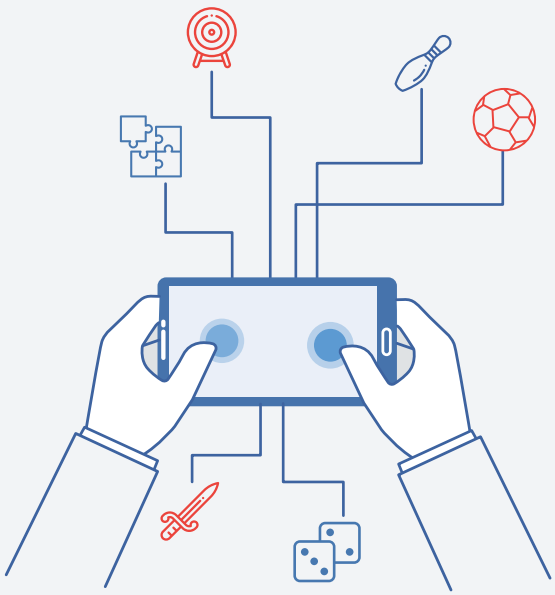
While game localization is not new (and it is indeed the norm in East Asian countries and most of Europe), game localization in Indian languages has not been given adequate attention.

The Government of India has also realized the necessity of localization and has mandated Indian language support on mobile devices.

The question isn't if there's any demand for content in one's own mother tongue or not. *The question is why mobile gaming in India hasn't been adequately localized yet.*

Unlike news, media and entertainment, or social media apps, gaming apps rely heavily on image frames, sound files, game physics, camera angles, and loads of visual, textual, and audio content to create an immersive experience.

Our report shows that only one game featured in the top 25 apps. Cross platform game engines only have limited support for Indian languages. ***Localization can fix this gap.***



Unity & Indic Fonts

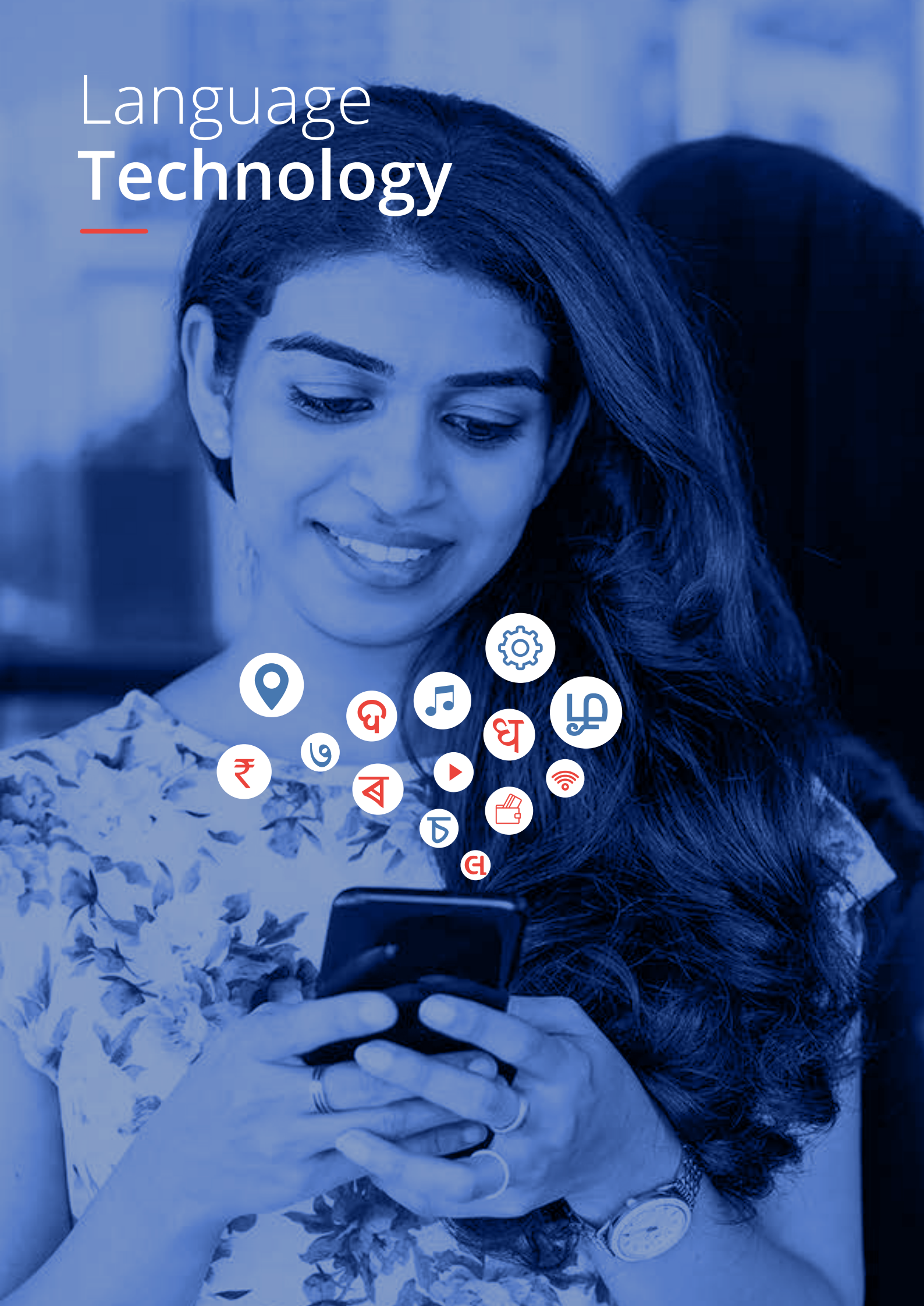
In any game, text is used throughout, from title sequences, to character selection, to instruction, to narration to credits.

While Unity supports development of games in both 2D and 3D, the way the text is defined and used in these separate interfaces is different. When it comes to Indian scripts, their inherent non-linearity – factors like correct mātrā (*Indic vowel diacritic*) positioning and conjunct formation – plays a large role. Proper text support is the cornerstone of proper Unity support.

By ensuring accurate text rendering for Indic scripts, with highly legible fonts, Unity developers can make sure that their games maintain a high level of playability across different devices, resolutions and text sizes.

How playable a game is, is directly tied to the experience it provides. Seamless, accurate text rendering provides an unbroken, seamless experience for gamers, making it possible to target India's 300 million Indian language users with games that they can play in their own language.

Language Technology



Mobile devices must come with basic support for Indian languages.
Fonts & display solutions must provide seamless visual experiences to users.

The Government of India's language mandate for mobile devices has ensured that all phones sold in the country from Feb 2018 onwards must support text in all 22 scheduled Indian languages, and input in at least two of them.

Content Consumption

A user's journey begins with them consuming content available on various platforms.

Content consumption, as a passive stage, requires localization services that deal with interface and content conversion. Content and platforms created in English can be localized for Indian language users through converting text strings into different languages. The 2 major tools for this are ***translation*** and ***transliteration***, which work together to provide users with content and interfaces in their language.

Content conversion mostly deals with text strings, and these strings can either be static, or unchanging, or dynamic, with updates featuring strings that are generated in real time, meaning that their content needs to be localized in real time too. ***The end result is that a user is able to consume content with no breaks in language experience.***

Engagement

Engagement requires more work on the part of the system, as it involves more activity from the user. This begins right from enabling input in Indian languages through typing and voice. Allowing users to interface with platforms in their own language ensures that these users can now engage with content and platforms.

In addition, cross language Indic search can help users look for and discover new content, including content that is in a different language than the search query. This helps increase the extent to which a user is part of the content ecosystem. Tools like Indian language chatbots, voice assistants, and more help extend the range of options available to an Indian language internet user even further.

Transaction & Service

To truly tap into the purchasing power of the massive Indian language user base, platforms must allow users to actively transact and use services in their own language. For example, wallet apps are more popular than bank apps, and many leading wallet apps are localized.

Transaction & services require additional layers of language support. The trust barrier that keeps users from transacting and using services online can be broken by ensuring a smooth transition to secure, localized pages. This flow can involve multiple platforms - a transaction page could involve a main site, a payment gateway, and a netbanking page. All these need to be localized.

Emerging Tech

Machine Translation

Machine learning helps power more accurate, precise automatic translation, something that's essential for localizing content at scale. It allows translation systems to learn from millions of examples and patterns and continuously improve the naturalness of its translation.

Indian languages have certain linguistic quirks that can confuse translation systems, like cultural differences, and stark differences in formal and colloquial vocab. Water, for example, can be jal or pānī depending on formality, and the wrong variant would sound horribly out of place. A machine learning system should be sensitive to these differences.

Intent & Sentiment Analysis

Analyzing user behavior and response to content helps companies create better content and target it better for increased positive engagement and reception. Intent refers to how a user could potentially engage with content or a platform, like a booking or a subscription. Sentiment refers to what they feel about a certain piece of content, like a status update.

As more Indians come online and become mature users, it will become increasingly more important to learn more about their usage patterns and how they react to content, in order to build better user experiences for them.

Voice

Voice search lets users find content by allowing them to speak to their devices. Indians who are coming online for the first time may be more comfortable searching by voice than typing, since Indic language typing would be something completely new to them. Voice, on the other hand, isn't. According to Google's own data, ***28 percent of Google searches done in India are powered by voice queries.***

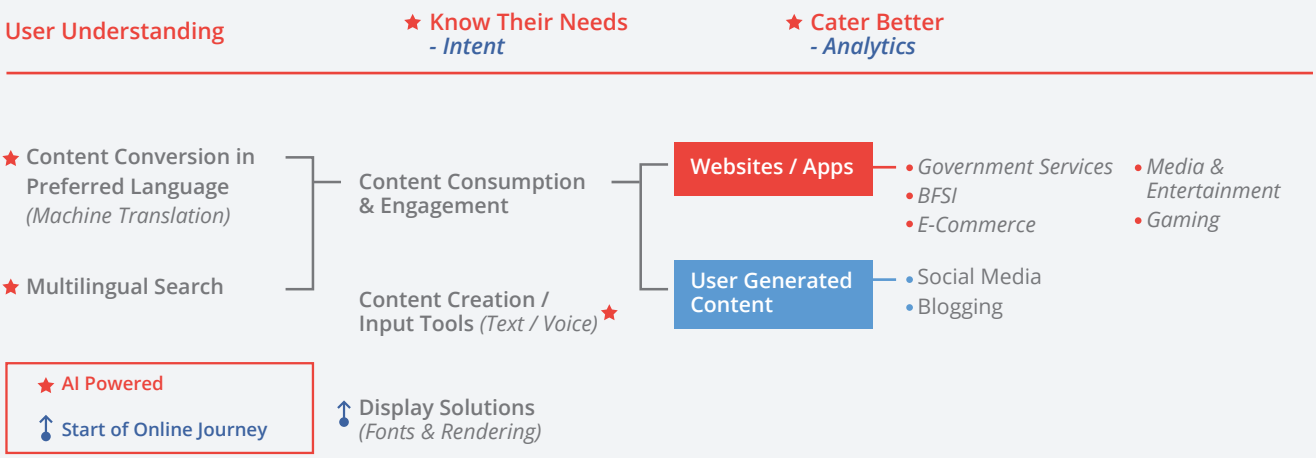
The rise of voice search represents a new set of possibilities for Indian language users, and with this user base's well-documented tendency to leapfrog, voice based services in Indian languages could possibly empower hundreds of millions of Indian through voice assistants and similar apps.

Indic Search

Indic search lets Indian language internet users find and discover content in their own language with greater ease. It widens the range of options available to the Indian language user and expands their available content ecosystem. Indic search involves both translation & transliteration.

Search should be able to navigate this multilingual landscape and index results in different languages and scripts, giving users a great degree of interoperability and flexibility across these different languages and scripts.

Role of AI in a User's Online Journey



AI requires large amounts of data in Indian languages in order to train systems & improve accuracy.
It involves parallel corpora, data sets that include data in the source language and data in the target language.

The Challenges

There's a very real scarcity in actual resources for building digital support for Indian languages. The European Union, for example, has EuroParl, a database of corpora (*parallel language vocab data*) for multiple European languages. Indian languages have nothing comparable. This means these resources for Indian languages have to be built from the ground up.

To support a large user base with seamless Indian language experiences, solutions and databases need to be built at scale, ***capable of supporting an entire ecosystem.***

The Experts Speak

This is a great report by Reverie and contains a lot of data-backed analysis and trends that have never been available in this market with respect to the behavior of Indian language internet users. The data and insights pertaining to the '*Engagement Levels*', '*Demographic Breakup*', '*Top App Categories*' and most importantly, '*The Evolution Path*' of Indian language internet user are quite valuable.

I believe this report could be really useful for a lot of Product Managers and Marketers in various companies across industry verticals.

Jayanth Kolla
Founder & Partner, Convergence Catalyst

The Business Case for Indian Language Experiences

Rewind three years and it would have been difficult to convince most businesses that offerings in local languages have to be a key part of a company's strategy. In early 2017, Jio came and changed that perception in a matter of months – suddenly 'Bharat' came online, uncapped and lit up this opportunity. Data prices have fallen from Rs.152 per GB in 2016 to less than Rs.10 per GB today making the internet affordable and giving access to a segment of society which is experiencing the digital revolution for the first time.

Media in India is predominantly consumed in local languages: for the week of 14th April 2018: the top 5 Hindi Entertainment TV channels saw 3 billion impressions vs. 1.1 million impressions for the top 5 English Entertainment channels (*source: BARC*). Similarly in print media: local language publications comprise of 87% of overall print circulation (*source: RNI, 2015*).

Despite this high engagement and reach on vernacular media, I still get resistance from the ecosystem of entrepreneurs and investors around value that can be created in building businesses for the local language user – *"these are lower income users!", "Will they pay?", "Local language advertising rates are low!"* and so on.

Yet the reality is that examples abound of valuable businesses created on the local language user: whether it's microfinance companies like ***Bandhan Bank*** which is the 8th most valued bank started as a microfinance organization, local language newspapers like ***Dainik Bhaskar*** with its 70 million readership and valued at $800 million or local television networks like South Indian languages-based ***Sun Networks*** valued at $5 billion.

Even on the digital front: we're already seeing emerging platforms create value. From ***Hotstar*** which generated $50M of revenue in FY17 to ***YouTube*** which reportedly is at an estimated $130-$150M of revenue, as investors we expect this is just the beginning of a long term cycle of value creation. While an individual user's inherent value or propensity to pay may be low, the 500 million users now accessible allows new kinds of business models which are driven by low-ticket size but multiplied through sheer scale.

There are few opportunities when entrepreneurs have the privilege to build products that can influence a user's first interaction with technology. Moreover I expect these innovative early models which will have a profound impact not just in India but even in other emerging markets seeing similar technology adoption waves.

Akshay Bhushan
Partner, Lightspeed India Partners

With over 8 lakh driver partners on our platform across 110 cities, Ola understood the need to develop local language app interface for its driver partners for them to function and serve the customers with utmost efficiency. Hence, our driver app supports 7 Indian languages apart from English to enhance the user friendliness of the app and help our driver partners and customers interact seamlessly.

*Startups like **Reverie** are betting big on this segment with regional language assistance. From a complete **LaaS platform** that supports languages like Hindi, Marathi and Tamil to building intuitive keyboards and a repository of apps, Reverie is playing a pivotal role to bridge the rural-urban digital divide; and has proven very useful in our efforts to build a robust driver partner ecosystem.*

Sumit Kumar
Supply Product Head, Ola (ANI Technologies Pvt. Ltd.)

With over 22 official languages and 1600+ Indic languages, content consumption amongst Indian online user is diverse. The majority of the Indian audience fails to explore the vast opportunity that the Internet offers because of limited language/content support online.

To get the next billion users online, the need of the hour is to create a favourable ecosystem for non-English users, by offering them a platform to create relevant and localised content for consumption in their local languages and ensure a seamless first time Internet experience.

Farid Ahsan
Co-founder & COO, ShareChat

REVERIE
Language Technologies

About Reverie

Reverie's vision is to promote Language Equality on the Internet™. It was founded in 2009 by **Arvind Pani**, **Vivekanand Pani**, and **S.K. Mohanty**, and is led by a strong R&D technology team that builds technologies to bridge the language divide in the digital world. Reverie works with OEM and chipset manufacturers, Central & State Government bodies, enterprise companies across industries such as the consumer internet (*online retail, e-commerce marketplaces, travel, media & entertainment*), banks and financial services, Government affiliated agencies, and the developer community.

Reverie's Language-as-a-Service™ (**LaaS™**) cloud platform provides localization services such as local-language translation, transliteration, device input, and search, through a set of APIs.

The platform integrates with business infrastructure (*websites and apps*) to enable end users to consume digital content in their preferred language, all in real-time.

Reverie was the only language technology company that participated in the NITI Aayog led **Champions of Change** initiative in 2017.

In Jan 2018, **Reverie** announced its **#PledgeToTeach** initiative, started with the intention to educate Indian language literate Indians about Indic typing, or typing in Indian languages. Internet users were encouraged to teach people how to type in their own language and access the internet as well.
Write to us at mrkt@reverieinc.com for feedback and queries.

Awards and Recognition

Qualcomm Ventures
Q Prize
TECH SPARKS
Microsoft
CODE for HONOR

Customers That Trust Reverie

eNAM
National Agriculture Market
Government of Karnataka
Government of Rajasthan
BHIM
OLA
vodafone
विजया बैंक
VIJAYA BANK
A friend you can bank upon
Delhivery
:practo
Qualcomm
Karbonn
LAVA
INTEX
XIOX Mobiles

REVERIE
Language Technologies

www.reverieinc.com