

# language

This document is an exclusionary act. Conceived and written in English by highly literate speakers of English, the world's most dominant and well-represented system of knowledge. But Wikimedia's vision aims to be *for all people, in their language*, calling for us to break down the exclusionary systems of knowledge creation and access. To do this, our systems, software and governance will have to evolve away from an English first, others maybe model, to a system that not only supports all languages but empowers them. For 60% of the world [1] using multiple languages, for different purposes, is a the normal way to live. Pidgins and inter-languages produce knowledge too, so how must the current one-project-one-language model adapt to support evolving languages such as Hinglish and Spanglish? And while capturing and growing knowledge in all languages is an uncontroversial goal, it is at these fuzzy edges that we can see, for example, the unintended exclusionary potential of technologies such as machine translation. As isolated cultural "space" collapses, a tension arises between support for cultural uniqueness and support for the experience of living between languages. And finally, language acquisition and cultural adaptation will be a huge factor in the foundation's success – what new capabilities does this demand of our organization and of the movement as a whole?

## Sections

- Privilege
- Where to Focus?
- Priorities
- Interoperability
  - Multilingual use patterns
- Support
- Opportunities & Threats
- Advocacy
  - Otherness
  - Preservation
  - Societal Goals

- Notes
- Sources

### Privilege

[not edited] This section is about the need to see and address our English first model. How does culture (both Foundation/Wiki community culture and the cultures of the underserved language communities) play into what languages and/or topics are “marginalised”?

- > Isolated cultural spaces are collapsing thus more languages are going to co-exist in the same space
- > Diversity of cultures needs to be better handled.
- > There will be predominance of some languages due to various factors like demographic majority, regional prevalence, social prestige, or academic needs
- > We will need to find common ground that can act as the acceptable connecting points to diversify from.
- > This will need more case based exploration. Language may not be the connection in all cases, but can be a catalyst.

### Where to Focus?

[not edited] How Do We Determine The Scope Of Influence Related To Languages That We Need To Consider?

Language is a vast source of information and influence. Small and large impacts arising from languages are always happening.

Specialized tracking of all kinds of changes related to languages will be a

heavy investment, without many clear benefits.

For the strategy plans, we need to isolate a few areas of impact that we expect to spill over into our work that we can build upon for technical and social intervention.

As language is as good as the living entities that define them, we need to track some of these defining factors and the changes that matter to them.

### Priorities

[not edited]How do language interventions get prioritized? What should drive this?

What Are The Global Factors That May Influence Language Trends For The Strategy Planning Period?

Since Languages are influenced by factors affecting the people themselves, what are some of the sharp changes we are already seeing around us:

- > Migrations: Forced migrations from conflict zones, Economic migrations for subsistence and specialized jobs
- > Cultural preservation
  - Increased attempts especially at places experiencing unusual amounts of sudden cultural intermingling
- > Increased purchasing power in emerging markets around the world
  - Improvement in lifestyle
- > Better access to education
- > More freedom in choice
- > Lower barriers+cost to adoption of technology

- > Cheaper devices and services
- > Increased digital literacy
- > Ease of use
- > Need for increased connectivity

Tracking cultural shifts due to economic migration and other current global factors is important to understand the adaptive linguistic uses that can impact the technology we operate with.

### Are Individual Languages Important In The High Level Strategy Process?

Yes. When we talk about languages in general, we also talk about the regions and people that we connect with them. By that extension, if we try tracking through factors like:

- > emerging markets for increased technology adoption
- > languages with need of additional content
- > languages with need of special support: this could also include the cases for additional multilingual support

The actual set of languages may not be restrictive, even if special focus is put on them, as resolutions for one or more languages may extend to others, and can be used for quick gains.

### Interoperability

Multilingual use patterns

[not edited] This section is about how the platform must support context switching

between languages (according to context of use).

(How) Can we create content that is consumable in multiple languages, or doesn't require language at all to enjoy and use?

The model of language around which the projects are built, with parallel encyclopedias and dictionaries for each language, and some set of shared content that is available across them all, understates the complex way people use languages. For 60% of the world (<http://ilanguages.org/bilingual.php>) using multiple languages, for different purposes, is a the normal way to live.

Code switch, pidgins and inter-languages produce knowledge too. What does our model of 1 project:1 language adapt to a world of Hinglish and Spanglish?

Language and culture are intimately related, in multilingual societies languages come to serve specialized purposes, and some types of knowledge are associated with a particular language. For example it is not uncommon for there to be a language for science, a language for government and a day-to-day spoken language, in which mass media and popular culture are conducted. The idea that each language will have a complete vocabulary and source material to build an encyclopedia that looks like English's doesn't stand scrutiny.

New formats needed?

## Support

[not edited] How do we provide not just content, but also support in marginalised or underserved languages? What happens when a contributor who only speaks e.g., Yoruba or Tibetan has a policy question? Participation demands support, especially on the long tail.

- Languages new to the internet, we need to continue pushing forward Unicode font support and basic text entry (reference-- Minh, Odia, etc), to ensure the most basic forms of access for all languages

## Opportunities & Threats

[not edited] This section is about threats to culture posed by machine learning and language extinction. How can we avoid reinforcing language inequalities if we use technology (machine translation, for example) to fill our content gaps?

Content translation is crucial, but insufficient. Language and culture are deeply bound up with categorization, notability and content expectations. Simply translating all articles between all pairs of languages would not make a relevant or even comprehensible experience for most humans.

Machine translation is making the boundaries between languages softer. Wikidata and Commons doing the same. How do we preserve the cultural values and self-determination of each speaker community in a world where these

communities and languages are becoming more porous and mixing more than ever?

## Advocacy

Otherness

[not edited] This section is about how or if we take on the role of extracting and cataloguing knowledge from languages that don't have a writing system? Our intervention could be at the level of platform - a DIY kit for capture? The hawaiian Oli.[x]

*“Since the beginning of time, every culture has developed means of passing on important information to its people. For Hawaiians, there was no written language per se until the 1820s. The missionaries introduced the alphabet which made it possible to represent Hawaiian language in the written form. Until then, all information was passed orally through the use of songs, chants, and poems.*

*Hawaiians devised various methods of recording information for the purpose of passing it on from one generation to the next. The oli was one such method. Elaborate chants were composed to record important information, e.g. births, deaths, triumphs, losses, good times and bad.”*

Preservation

This section is about the extent to which the WMF should advocate for or orchestrate the preservation of endangered languages. Is partnership the way to go here? Developing educational materials in languages that are spoken by

---

# Culture

05

many people and aren't immediately endangered, but are rarely used for education and business

## Societal Goals

How can the Wiki movement contribute to larger societal goals? Should it?  
Facilitating language teaching and learning

---

## Notes

[1] (<http://ilanguages.org/bilingual.php>)

[2] Hawaiian Oli

<https://apps.ksbe.edu/olelo/learning-place/performance-indicators/chant/foundational/significance-oli-chants-in-hawaiian-society>

## Sources

R. Bhattacharjee and C. Gauthier [Research and Insights](#), Other contributors<sup>1</sup>: A. Amaroni, D. Horn, R. Isler, J. Katz, J. Minor, T. Negrin, M. Novotny, N. Pangarkar

<https://annual.wikimedia.org/2016/fact-1.html>

Google-KPMG report titled 'Indian Languages - defining India's internet

<https://assets.kpmg.com/content/dam/kpmg/in/pdf/2017/04/Indian-languages-Defining-Indias-Internet.pdf>

The Rise and Rise of Hinglish in India

<https://theconversation.com/the-rise-and-rise-of-hinglish-in-india-53476>

→ Article coverage by language

- ◆ Research: [Increasing article coverage](#)
- ◆ [Growing Wikipedia across languages](#)

---

<sup>1</sup> If your name was left off the list by mistake please contact JMinor or MNovotny

- ◆ [Content translation recommendation API](#)
- ◆ [Expanding articles cross-language](#) @ Meta
- New Voices synthesis: [cultural issues hindering content creation](#)
- Language Extinction:
  - ◆ [Wikitongues](#)
  - ◆ [Foundation for Endangered Languages](#)
  - ◆ [Rosetta Project](#)
  - ◆ [Living Tongues Institute](#)
  - ◆ [UNESCO linguistic diversity on the internet project](#)
    - And [an associated report](#)
  - ◆ BBC: [Why we must save dying languages](#)
  - ◆ Book: [Linguistic Genocide in Education](#) - Tove Skutnabb-Kangas