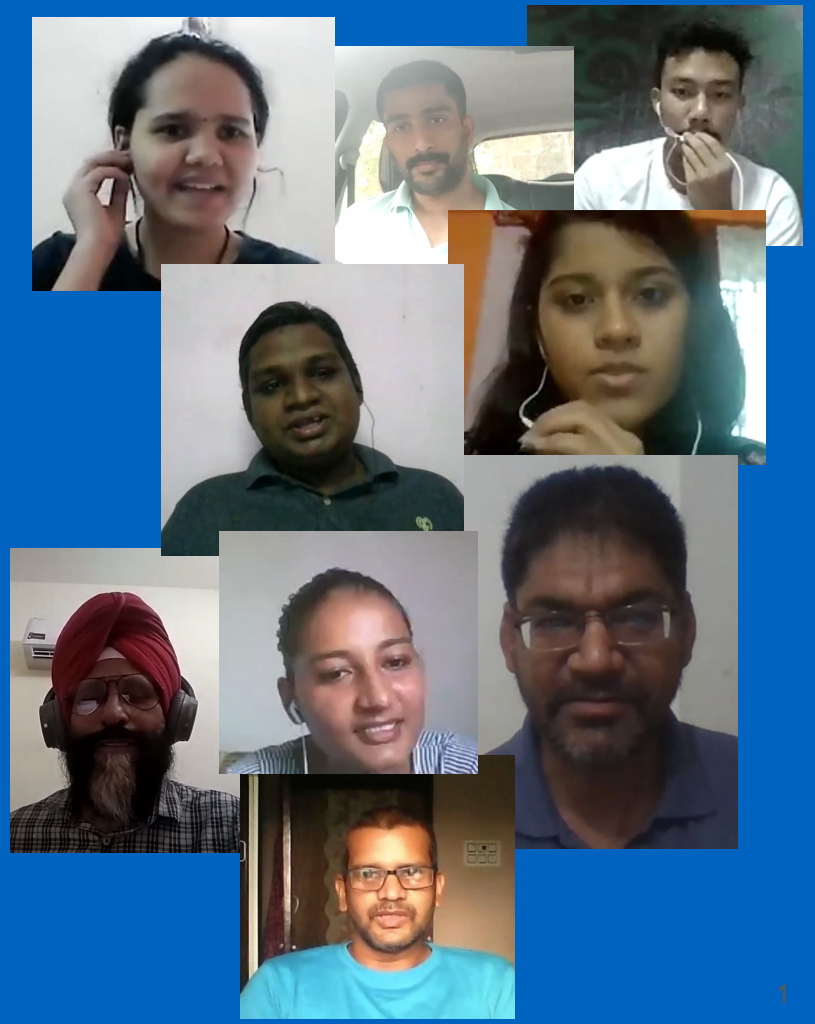
Multilingual Editors' Experiences Research

Product | Design Strategy | Language Team

Prepared by Eli Asikin-Garmager
November 2020



WIKIMEDIA
FOUNDATION

Supporting multilingual editors in **small wikis** who are leveraging **translation** to contribute across knowledge and **content gaps**

WIKIMEDIA
FOUNDATION

Table of Contents

1. Introduction

Problem statement & overview

Research goals & questions

The communities/wikis

Executive summary

Problem Statement & Overview

Worldwide, often countries with the fastest growing online populations are home to languages with relatively small Wikipedias (wikis).[1]

Thus, if the Wikimedia Foundation will achieve its goal of removing barriers to knowledge, it must address language barriers and the socio-cultural needs of these communities, including many in South and Southeast Asia.

Despite being small relative to larger wikis, many small wikis have significant growth potential based on language community size and growing number of active editors.[2]

Adding to this potential of increased knowledge access, previous research has indicated that multilingual editors may be more active than their 'single edition (monolingual)' counterparts.[3] This suggests that these editors are important actors for helping to close content gaps, or areas of knowledge that are only accessible on certain language wikis but not others.[4] In addition to the benefit of making knowledge more accessible in these communities, their members are also able to contribute local knowledge, otherwise not available to the global community. This content could then just as easily be translated into the languages of larger wikis, resulting in mutual benefit. In this way translation can also help counter colonial patterns of knowledge distribution.

Supporting multilingual editors in small wikis who are leveraging translation to contribute across knowledge and content gaps

An important way of addressing language barriers and content gaps is through translation. The Wikimedia Foundation's Language Team's Content Translation tool has been used to create over 750,000 articles since its inception.[5]

As part of its annual plan, the Language Team aims to help small wikis grow with translation and improve the process of creating multilingual content. In order to reach this goal, the team aims to better understand the needs of small wikis and their editors. This includes identifying barriers and challenges to the use of translation and creation of new content. To help make multilingual editors more productive in creating new content, **this generative research project investigated the current experiences of small wiki editors, and aimed to better understand the workflows of potential editors who currently create online content with the aid of translation.** For both current and potential small wiki contributors, the current project focused on workflows, Wikipedia experience journeys, motivations, needs, and barriers to editing.

[1] To explore these trends, for online population trends see: https://www.statista.com/statistics/292488/fastest-growing-internet-populations/; for current language Wikipedia sizes see: https://meta.wikimedia.org/wiki/List_of_Wikipedias

[2] For strategy purposes, in their annual plan, the WMF Language Team has defined "small" as wikis with fewer than 100k articles and less than 100 translations per month. The threshold of 70 editors (and >20 active editors) per month is hypothesized as important figures for potential growth.

[3] https://arxiv.org/pdf/1312.0976.pdf

[4] For a definition of content gap see: https://meta.wikimedia.org/wiki/Research:Content_gaps_on_Wikipedia; and for more details on what we know about current content/knowledge gaps, start with: https://meta.wikimedia.org/wiki/File:Knowledge_Gaps_%E2%80%93_Wikimedia_Research_2030.pdf

[5] https://en.wikipedia.org/wiki/Special:ContentTranslationStats

Research Goals

The primary overarching research goal of this generative research project was to **better understand *small wiki*, multilingual editor experiences** — particularly those involving translation. More specifically, we wanted to increase our knowledge of small wiki multilingual editor experiences - both those of current and potential contributors.

This involved investigating the socio-cultural context of the target communities, and identifying specific language-related considerations and concerns. These language factors included structural, written, and sociolinguistic factors. Key sub areas of focus included editor values and motivations, barriers to content creation, and awareness/knowledge of the Wikipedia ecosystem, especially translation and multilingual features.

Using phenomenological methods, **the project emphasized task-based inquiry to uncover content creation workflows and provide evaluation of current and potential contributor experiences with current Wikipedia translation tools.**

Research planning also involved a desk review of relevant previous research on translation and editing. An annotated bibliography on 'Multilingual Editors and Translation' was prepared to summarize a selection of relevant articles and readings.

Primary research questions

1. How do small wiki contributors edit and create content? What are typical content creation and translation workflows? To what degree (and how) are these workflows exclusively supported by Wikipedia, or not?
2. How do editors discover and experience Content Translation? What are common entry points, discoverability problems, usability issues, and barriers/motivations to use?
3. What are multilingual editors' motivations to use translation, and what Content Translation value propositions resonate most?
4. For design purposes, what are small wiki editor personas?
5. What barriers do editors face when using Content Translation? Why do some continue to use it while others stop?
6. How do users understand and think about translation, especially as it relates to Wikipedia, machine translation (MT), and use of MT in tools?

The Communities/wikis

The team used a number of criteria to identify which small wikis with potential for significant growth to focus on for this project.*

In arriving at this set of languages, we considered metrics such as *speaker population size, number of active editors, number of translations, mobile views*, as well as *general feasibility.*

Note: At the time of research, Content Translation was out of beta for this set of wikis.

Language	Translations	Articles	Users	Active editors/mo	MT	Mobile views	Speakers
Telugu	1.5k	70k	95k	49 (+10/mo)	✓	80%	93M
Malayalam	7.6k	68k	135k	90 (+15k/mo)	✓	76%	38M
Kannada	1.4k	26k	62k	37 (+4/mo)	✓	80%	56M
Bengali	7.6k	84k	273k	266 (+75/mo)	✓	86%	265M
Gujarati	47	29k	56k	16 (+3/mo)	✓	75%	61M
Marathi	385	57k	120k	69 (+23/mo)	✓	87%	95M
Punjabi	8.3k	34k	32k	35 (+6/mo)	✓	50%	116M

*Please see appendix for a full discussion of the rationale behind the selection of these 7 wikis. The appendix also contains details on how these statistics were gathered and compiled.

Executive Summary - *Top 10 Key Takeaways*

1. Most participants, including both current and potential editors, were unable to successfully navigate to Content Translation (CX), despite consistently searching in many of the same places. Many pre-translation workflows begin at large wikis where editors are reading.
2. Current editors and newcomers are delighted by a number of CX features that speed up the editing process and require less typing, a challenge for some users. Delighters included automatic links and references, side-by-side presentation of the source/target article, and help generating key article vocabulary.
3. Keyboard input challenges with non-Latin scripts are widespread. This is not a CX-specific problem, but impacts the CX experience in critical ways. While some current editors have mastered non-Latin inputs on WP, others have developed a range of workaround solutions. Without the ability to type fluently, users are unable to edit machine translation outputs.
4. Observation of editing workflows revealed that 17 of 21 current editors sampled relied on external tools for at least one of the following purposes: input, drafting, translation. These external supports are helping editors accomplish goals such as entering non-Latin scripts, drafting new articles, and translating article sections.
5. Alerts/warnings in Content Translation are not always well-understood and timing of their presentation is critical. Editors didn't always understand how to to resolve alerts, and not clear on how to proceed if ready to publish but not able to due to a machine translation use limits.
6. An editor's motivation is multi-dimensional. Motivations range from purely altruistic in nature to those grounded in self-improvement. During research sessions, we observed 6 primary patterns: improving access, expanding contents, self-improvement, social, audience, recognition/identity. (more details, slide 24)
7. Small wikis present unique environments for editors compared to larger wikis. For example, editors noted that when compared to the English wiki, there's often much less anonymity and more receptiveness to new editors. On the other hand there are fewer editors to complete admin tasks, proofread articles, and help find sources.
8. In most cases, participants faced problems related to a specific task. They wanted to troubleshoot a very specific, contextual problem, not access a lengthy help page. They're looking for contextual help, and new translators need help building the support systems that experienced translators report having for help and motivation.
9. Small wiki editors vary in terms of whether they are primarily mobile-first editors or rely on laptops/chromebooks. In general, editing and reading activities correlate with device use patterns. In general, we found that input of non-Latin scripts is slightly easier for potential editors on mobile devices.
10. Despite nearly 100% brand awareness, the editing process remains below the tip of the Wikipedia iceberg. Multilingual readers often default to reading on larger wikis, and translation as a form of editing lacks widespread awareness - even among current editors. Social factors are in part determining how editors contribute and what topics they contribute about.

2. Research Approach

Participants

Methodology

Data collection process

WIKIMEDIA
FOUNDATION

Participants

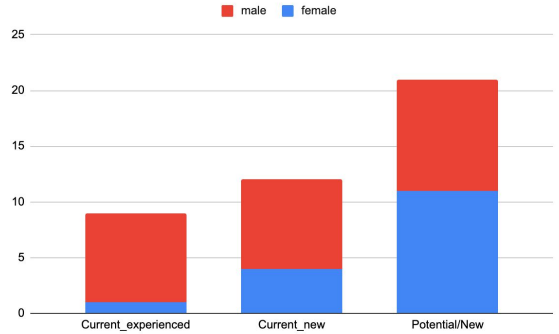
Overview of participants

- **42** participants across **7 wikis/language groups**:
  - **Marathi** (mr)**, Punjabi** (pa)**, Kannada** (kn)**, Bengali** (bn)**, Malayalam** (ml)**, Gujarati** (gu)**, Telugu** (te)
  - 21 current editors - 9 experienced / 12 new (new=account created in 2017 or after)
  - 21 potential editors
- **Professions/areas of study** - advertising, business, engineering, computer science, education, journalism, medicine, science, technology
- **Age** -
  - Current editors: *18-25*(1) *26-34*(9) *35-50*(8) *51+*(3)
  - Potential editors: *18-25*(3) *26-34*(6) *35-50*(12)

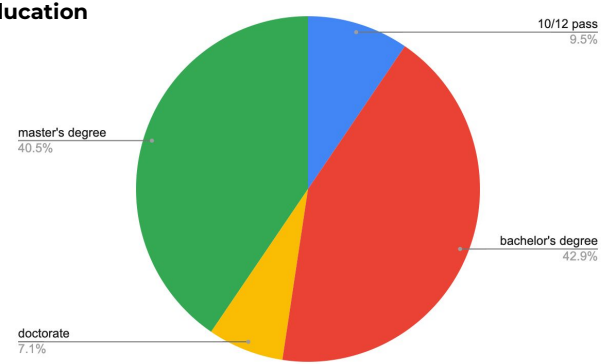
Participant recruitment process

- Community pump announcements combined with direct recruitment messages to contributor talk pages, all in local languages
- Potential participants identified through Quarry[7] and contacted via talk pages
- Recruitment of potential editors in collaboration with Anagram Research
- Interested participants responded by completing a Google Form screener, and the researcher followed up with scheduling options

Gender



Education



[7] https://quarry.wmflabs.org/

Methodology

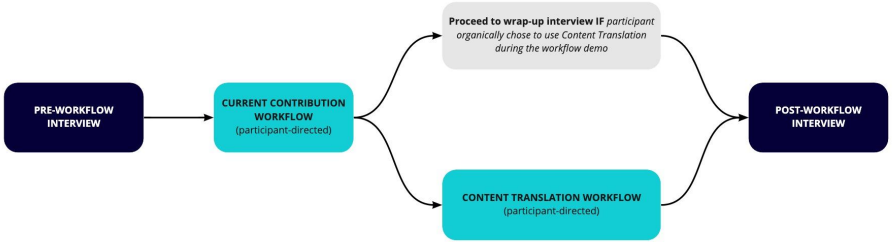
Method & approach

- Semi-structured contextual inquiries, involving a combination of workflow observations and interviews (overview of research sessions shown on right)
- Research Session Protocols: current editors; potential editors
- Progress updates/executive summaries/video clips delivered to design & language team along the way

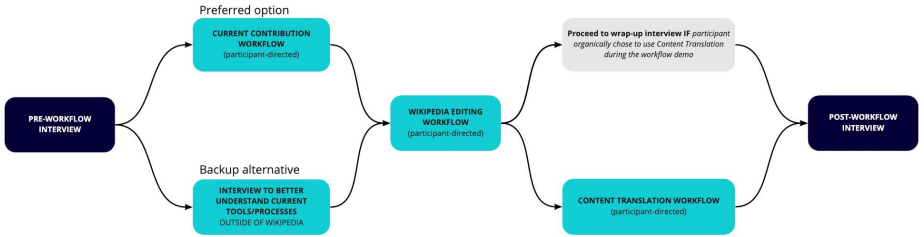
Research sessions

- Moderated, remote sessions using Zoom
- Conducted in language of participant's choice (8 different languages)
- Mix of mobile-only, mobile/laptop, and laptop-only sessions (device choice determined by participant)
- ~2-2.5hr session duration
- All sessions had audio, video(optional), and screenshare recorded, and were later transcribed and translated to English

Current editor research session flow



Potential editor research session flow



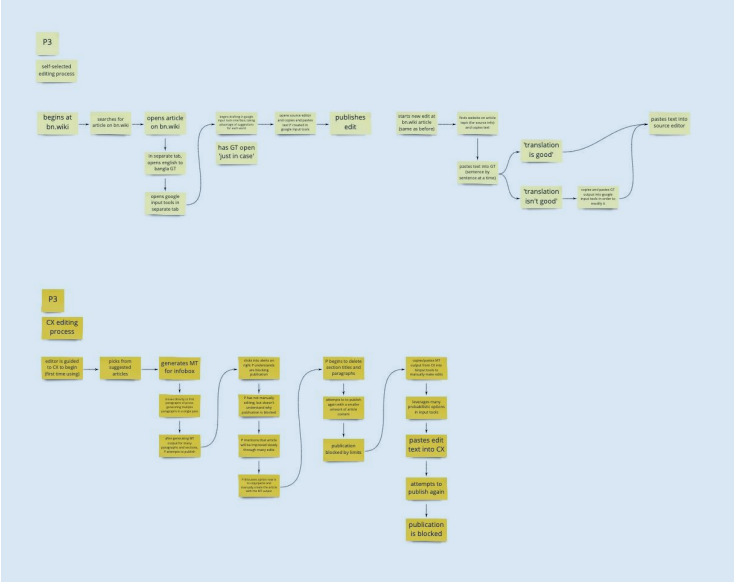
Data Collection Process

Session recordings → **Transcripts** → **Dataset**

With participant consent, a recording was made for each session's audio, screenshare, and video(optional).

Each recording was transcribed and translated to English.

A total of **1256** data points were extracted from the videos and transcripts and organized into a dataset for analysis and synthesis



Results of process

1. A dataset of 1256 data points from the sessions - data points coded & synthesized for this report
2. Inventory of external tools used by current editors
3. Maps of participant workflows, both while using Content Translation and editing Wikipedia more generally

3. Results

Content Translation

Discoverability

Moments of delight

Pain points & opportunities

External editing & translation tools

Alerts

WIKIMEDIA
FOUNDATION

🔎 Content Translation (CX) Discoverability

Most participants - both current and potential editors - were unable to navigate to CX

This included editors who had used CX previously (e.g., CE37_TE), and was observed across all 7 wikis.[10] The challenge was consistent, but so too were the locations participants were looking for CX.

Where are they searching for CX?

The left sidebar ('tools' or language options) was the most frequently searched location

The 'contributions' menu at the top of the page didn't receive many clicks because participants interpreted 'contributions' to mean previous/past editing activities, not an entry point for a *new* contribution (e.g., CE26_MR, CE27_MR, PE12_GU). When participants couldn't locate a translation option, they resorted to Google searches and/or Google Translate.

1. **Left sidebar** - 'tools' or language options
   [CE3_BN, CE25_MR, CE27_MR, CE33_PA, CE37_TE, PE18_KN, PE24_ML, PE36_PA, PE40_TE, PE12_KN]
2. **Top menu options** [PE6_BN, PE12_GU]
3. **Prompt to use CX from English Wikipedia** [CE7_GU, CE31_PA]
4. **Inside the editing interface options** (e.g., Visual Editor options) [PE12_GU, PE29_MR]
5. **Main settings** (a less common 'last-resort' move) [PE35]

For many current editors, the 'pre-translation/editing' workflow begins at a large wiki

Many participants began their editing activity by identifying an article of interest at a larger wiki (e.g., English wiki). After identifying a potential article, most participants tried to immediately determine if the article was already available on another wiki they contribute to.

For potential editors, UI translations sometimes contained 'uncommon' or 'unfamiliar' words, which added to the challenge of navigating Wikipedia menu options. Adding to the unfamiliarity of menu options, potential editors are often accustomed to navigating Wikipedia almost exclusively through a series of Google searches and search result pages.

🌱 OPPORTUNITIES

1. **More prominent CX entry points** to increase discovery rate. Leverage the locations participants are naturally looking for a translation tool, and consider making entry points more prominent when editors are reading on other wikis and considering articles to translate.
2. **Integrate translation (or minimally its entry points) with the main editing interfaces.** Despite being a relatively easy mode of editing for newcomers, CX is subsumed to other types of editing by being challenging to find and because it's embedded in menu options.
3. **Surface translation opportunities for editors contributing across different language versions of Wikipedia.** Potential CX users are missing out on WP-specific editing advantages of CX by falling back on other external, non-integrated translation support services, such as Google Translate.

[10] For all participant references and quotes, CE=current editor and PE=potential editor. Participant numbers are used to protect participant anonymity, and language codes include: MR=Marathi, PA=(Eastern)Punjabi, KN=Kannada, BN=Bengali, ML=Malayalam, GU=Gujarati, TE=Telugu. The identifiers are used for all direct quotes and in other cases as part of exemplatory (not always exhaustive) lists of pattern descriptions.

🔎 Where are editors searching for a translation tool option?

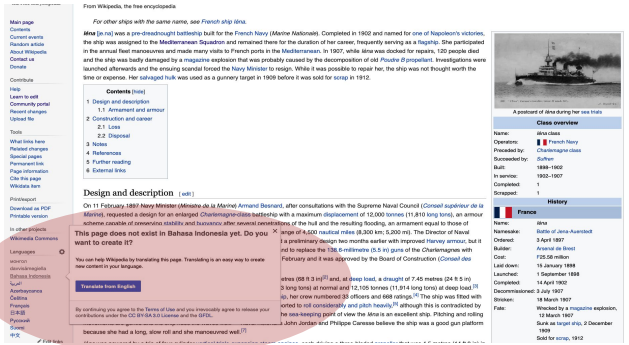
❗ **#2 Header menu options**

(...but not looking under “Contributions”)

❗ **#1 Left sidebar**

“Tools” →

Language options →

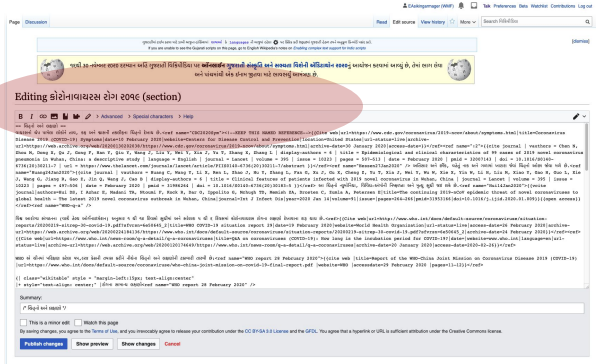


✔️ **#3 Prompt from another wiki**

(and *sometimes* finding it, depending on the article + language combination)

❗ **#4 Editing interface options →**

(Either visual editor or source editor (shown))



✨ Content Translation (CX) Delighters ✨

CX reduces the time demands of editing. Moreover, it provides scaffolding for many of Wikipedia's technical and conceptual challenges.

Both current and potential editors immediately note the time that CX saves them (e.g., CE15_KN, CE20_ML, CE7_GU, CE8_GU, CE31_PA, PE36_PA). Participants mentioned a number of different reasons for *why* CX is *faster* and *easier:*

- Notability is already established via the source wiki article
- Links and references are automatically inserted in the target article
- No coding is required to contribute
- Side-by-side presentation of source/target language
- Machine translation eases the burden of writing from scratch

For newer and less-experienced editors, these delighters are particularly critical as they reduce the technical and conceptual demands of contributing to Wikipedia. For example, not only must new editors learn a number of new interfaces and technical details, they must also learn cultural traditions of notability and reference-giving. Many of these challenges are automatically solved when using CX.

For all editors, as we'll observe later in this report, easing the burden of writing from scratch pays exponential dividends for editors facing text input challenges related to non-Latin scripts.

'Translation' encompasses many different mental models

While CX has many different features bringing value that users appreciate (such as those noted above), machine translation (MT) outputs are often the most salient - and memorable - aspect of the tool. In fact, it's not uncommon for "machine translation" to get used interchangeably with "content translation". Unfortunately, if a user doesn't anticipate needing to fix poor MT outputs, this unexpected burden quickly overshadows any delighters.

Less typing, less time

"In an hour your work is done. If you normally need a week, with CX you can finish in 2 hours"
[CE7_GU]

"When we do language translation, it becomes easy for us to make the article. My typing time is saved. I have to do only minor changes. When I was doing it manually I had to type 100% of it" [CE8_GU]

It's "easy"

"I don't have to collect the references. I just have to fix its translation or some spellings or grammar. So, it is much easier for me. And the other problem of notability also gets solved because the article is already available in English so it means it's notable" [CE9_GU]

A newcomer's perspective

"Content Translation turned out to be really helpful and overall, I get help instantaneously. For example, we can read line-by-line and we can understand. At the initial phase, I think this is very beneficial. As for the rest of editing, I will have to go back to Google Translator" [CE35_PA]

Language Input & Machine Translation
(CX Pain Points)

Keyboard input and typing tools present a persistent challenge for many participants

Input challenges were very widespread among potential editors, but also observed among current editors. Current editors had solved input challenges in a variety of ways. Some had mastered non-Latin script inputs on WP, but others drafted articles in Google Input Tools or Google Translate and copy/pasted back into WP (e.g., CE1_BN, CE2_BN, CE27_MR). An assortment of patchwork solutions underscores how widespread the challenge remains.

Input challenges are not a CX-specific problem, but impact the CX experience in critical ways. So, addressing this pain point affects how an editor will draft an article from scratch and what tools they'll use, but the challenges can serve as a total blocker for CX editing. Without the ability to edit MT outputs (due to input challenges), an editor is at risk of publishing a sub-par article, deleting problematic paragraphs of MT output, or simply abandoning the task altogether (e.g., PE23_ML, PE24_ML, CE27_MR). Without the ability to fluently edit MT outputs, the main workaround we observed was participants copy/pasting out of CX into outside tools, and then back into CX (e.g., CE33_PA, PE18_KN). It's not likely these editors would have always persisted in these activities outside of a research setting.

Machine translation (MT) outputs can overshadow other aspects of CX

Benefits of CX can get overshadowed by low quality MT outputs. If the MT output is gauged as too low quality, it is not uncommon for an editor to simply delete the selected passage and/or write it from scratch. Unfortunately, many participants didn't realize they could use CX without the MT outputs component and/or add new material to the target article that wasn't present in the source.

Even for participants who express displeasure with MT outputs, the vocabulary support ('saving time thinking of words I need') is often mentioned as very useful support. One of the problems with this vocabulary support is when an editor needs to repeatedly change the same word throughout an paragraph or article because the MT didn't accurately take context into account. Repeatedly correcting these contextual vocabulary choices becomes burdensome.[12]

[11] The same experience with MT outputs having serious issues was noted by a variety of participants, including CE26_MR, CE38_TE, CE7_GU, CE_8_GU, CE9_GU, CE31_PA. Participants varied in terms of whether or not they persisted or abandoned the articles.

[12] Research results provide support for prioritizing work such as: https://phabricator.wikimedia.org/T96165

Input challenges are widespread

"The first challenge is typing. Not for me particularly because I have been typing in Gujarati since 2016, but when I try to teach many people to edit Wikipedia the first challenge is typing." [CE9_GU]

There's a tension between the advantages of CX and the disadvantages of MT

"Sometimes the entire thing will be wrong. It will not be connected. [...] That is one reason I don't use it [CX]." [11] [CE19_ML]

"Yes, Google Translate (GT) is the only problem with Content Translation (CX), but if we don't use the GT [MT option in CX], we will miss the links" [PE17_KN]

"...you will not get anything else other than what is in that page. [...] I cannot include anything new there." [CE13_KN]

Collaboration & Mental Models
(CX Pain Points)

An unfinished translation introduces an undetected 'pause' to the system

After a translation is initiated, and until it's published, there's a pause in an otherwise continuous editing ecosystem because other editors may not know the translation has been started. Moreover, while non-translation editing is inherently collaborative on Wikipedia, CX nudges towards the long solo-task of creating a new article.[13]

There are missed collaboration opportunities and painful first experiences

Potential CX users sometimes don't know where to start, and where to go to get help (e.g., PE18_KN, PE28_MR, PE36_PA, PE24_ML, PE6_BN). More generally, participants expressed interest in translation being a more collaborative activity, especially when it comes to getting help translating difficult words and phrases, and ensuring consistency across articles and article sections.

While most new editors know there's a community of editors out there, unfortunately they sometimes don't learn where communication and socialization takes place until a negative message comes in. Moreover, given the atmosphere around translation in any particular wiki, feedback from other editors can be very different depending on if a new editor contributes via translation or by using Source or Visual Editor (e.g., CE38_TE).

A paragraph of text is easy, but other article elements present formidable challenges

Newer users often opted to begin translation at the first paragraph of prose, frequently skipping over article titles and infoboxes (e.g., PE23_ML, PE18_KN, PE42_TE). Even participants who moved fluently through translation hit roadblocks with templates, infoboxes, and tables. For example, one participant didn't even know what a template was, let alone what it meant that 'the template didn't exist in the target message' (per an error message received) (PE6_BN).

[13] While there is no 'minimum' amount of content required for article creation with CX, it was also not clear to participants when their article was ready for publication. Moreover, once published, there's currently no way to re-enter CX and use translation to add to the article. Section Translation will hopefully remove some of these limitations.

Collaboration and support is time-sensitive

"Another person may not know that someone has already started writing that article, and they would start it again" [CE39_TE]

"In the beginning, I didn't know that I could consult them [other editors]. I did a translation and everyone heavily criticized it, so I had some discussions and we came to an understanding. That was my first article, so I didn't know about it" [CE38_TE]

Knowing where and how to start once in CX

"I need someone to just guide me. Tell me what to click on to start" [PE18_KN]

"I've selected to translate, but I don't understand what to do next" [PE28_MR]

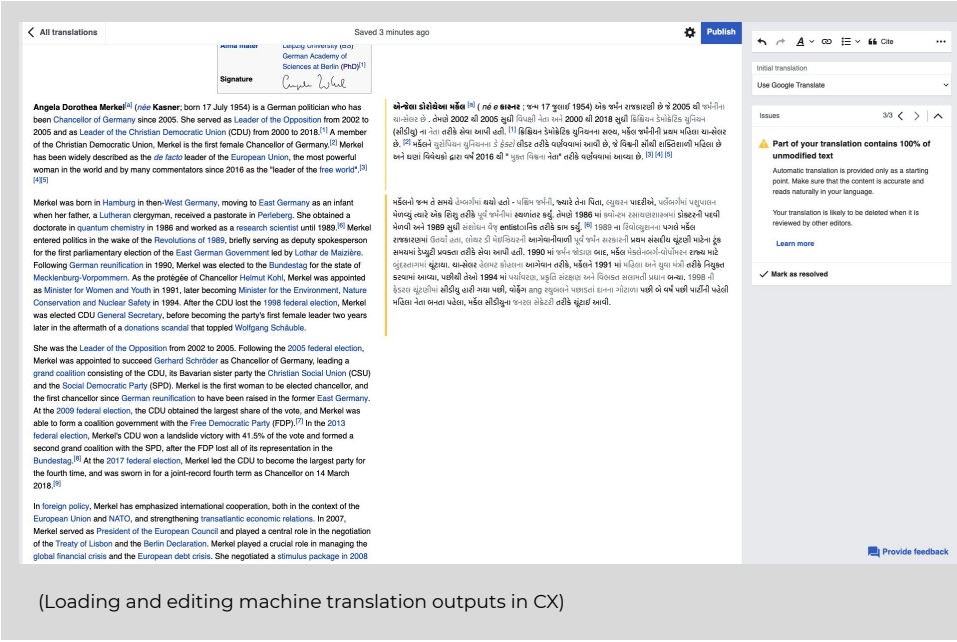
Setting up editors for success

"When we see a paragraph, psychologically we will feel it's a big paragraph to translate. That is very common, but here I feel that this is a very short sentence" [CE37_TE]

Supporting language input, machine translation interactions, & collaboration

OPPORTUNITIES

1. **Editors are in need of input supports for non-Latin scripts**, both in and outside of Content Translation. Any current keyboard supports should be made more salient and designed with existing expectations in mind (expectations driven by interactions with options like Google Input Tools and other tools providing predictive inputs.
2. **Build shared expectations around MT outputs and interactions.** Help editors understand how to interact with the MT. Increase awareness that 'CX without MT' is an option, and that editors may add novel content to target articles and sections.
3. **Smarter vocabulary support options.** From dictionary tools to section and article translation memory, providing vocabulary support will cater not only to the typical CX user, but those who are less interested in the sentence structure MT outputs.
4. **Revisit CX alerts and remove any jargon or content that assumes background knowledge**. And, if more context or explanation is needed, connect users with that knowledge through contextual help and/or contextual links.
5. **Allow collaborative translation of articles**, or minimally awareness to other editors if a translation is in progress.
6. **Explore ways of fosters 'translation support systems'**, and what it take to promote positive first social interactions for newcomers.



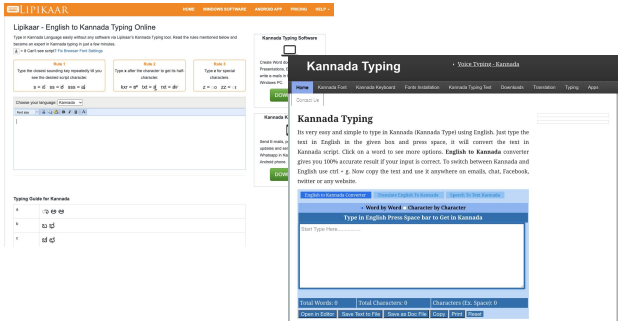
(Loading and editing machine translation outputs in CX)

External Editing & Translation Tools

During the research sessions, a portion of current editors only used Wikimedia tools to edit Wikipedia, aside from needs related to references and sources.

However, observation of editing workflows reveals that **17 out of 21 current editors relied on external tools for *at least one* of the following purposes: *input, drafting, translation.*** These 21 editors had a range of editing experience and tech savviness.

An inventory of the tools we observed being used to edit Wikipedia is shown to the right. Numbers in parentheses indicate the number of participants observed using the external tool. 'Input' captures any use of a tool primarily for the purpose of entering a non-Latin script. 'Drafting' captures activities such as writing sentences and paragraphs for the purpose of putting ideas together.



Translation	Drafting	Input
Google Translate - mostly sentences (5)	Google Input Tools (1)	Google Input Tools (2)
Google Translate - mostly words (4)	Google Keep (especially voice to text option) (1)	Google Translate (2)
Hard copy dictionaries (2)	Google Translate (3)	Other online input tools (5); for example: Lipikaar or Kannada typing
Other online dictionaries (1)	MS Word (2)	
Computer-aided translation tool (1)	Notepad (1)	
Google searches to distinguish word choices (1)	Other online input tools (2)	

⚠️ Content Translation Alerts

Alerts are not always well-understood

While some method of communicating critical information is required, when and how we communicate the messages may impact how they're received. In some cases, the message itself was unclear or misinterpreted. For example, a few participants interpreted the 'percentage of unmodified text' as 'percentage complete with the translation'. In other cases, keywords like 'unmodified text' were the root of confusion.

Timing is important

Because MT limit alerts are presented immediately, many participants' attention was drawn to alerts instead of editing the MT outputs, the latter of which eventually renders the alerts unnecessary. Moreover, upon beginning the translation task, both current and potential editors found early presentation of alerts stressful.

A clear call-to-action for how to resolve alerts

Finally, participants were unsure how to resolve some of the alerts. For non-blocking alerts, some participants clicked through the alert warning an article already exists in the wiki, not understanding the gravity of how their actions could overwrite the existing article, or what alternatives existed. For blocking alerts, participants faced the question: "If it's good, then what?", meaning what should an editor do if they believe their translation is of appropriate quality despite a warning that a sufficient amount of MT output text has not been modified.

🌱 OPPORTUNITIES

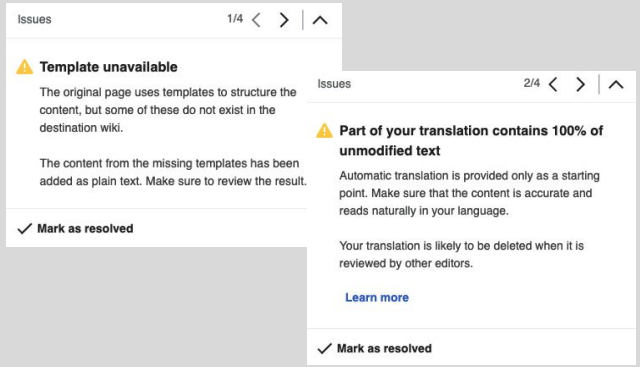
1. **Revisit the language of alerts to ensure each directs the user to a next step or resolution**. In some cases, contextual help or additional information may be needed (e.g., templates).
2. **Delay presentation of MT limits alerts** to encourage editors to focus on editing the MT outputs. Alternatively, present the alerts with different language when the user has not yet had a chance to modify MT outputs.
3. **Turn hard stops into possibilities for collaborative workflows.** For example, in place of blocking publication if the MT limit is not met, allow publication for peer review and feedback.
4. **Better understand the modifications that editors make to MT outputs** to allow possible improvement to the nature of limits and how they work. This proposal represents a possible first step.

> ***"We would get a repeated warning. It is very irritating"*** [CE38_TE]

> ***"We do translation. If it shows us an error, we get tensed. We don't know what to do"*** [CE27_MR]

> ***"For the overall ease of translation, I got errors, so it detracted from my translation experience"*** [PE12_GU]

> ***"I couldn't understand 'modified word'. I couldn't understand the meaning of 'too much unmodified text'. What is 'unmodified' in this text that's blocking publication?"*** [PE25_MR]



4. Results

Small wiki experiences

Motivation

Challenges

Where to get help?

Additional device patterns

Tip of the iceberg problem

WIKIMEDIA
FOUNDATION

Editor Motivations

An editor's motivation is multi-dimensional. Editors contribute to Wikipedia for a variety of reasons. While certain motivations may be stronger for one editor compared to another, all editors are motivated by a range of reasons. Motivations range from purely altruistic in nature to those grounded in self-improvement and personal interests. During research sessions, we observed 6 primary patterns of motivation, summarized below from the perspectives of editors.

Access for others

Not everyone can speak English or Hindi, so content in local languages is critical for our communities.

Wide reach

I can reach the greatest number of people on Wikipedia, making my work more valuable. Knowing how many people read articles I contribute to is highly motivating.

Self-improvement

Editing helps me learn new skills, from improving my language abilities to advancing my technical abilities. [15]

Recognition / Identity

Being recognized for my efforts makes me happy and ready to keep editing. My personal reputation as an editor is tied to my articles and their quality.

Social factors

Working collaboratively helps keep me motivated and gives me confidence that I'm editing correctly and my articles will reach readers.

Wiki Boost

I want to have an impact on the amount of content available in my wiki. It's important to know that my wiki's content is just as up-to-date as that of other, larger wikis. [14]

[14] Participants talked about the 'webs of content' that exist through linked articles. Editing may begin with one article, but then continue at the related/linked articles that may also not yet exist in a wiki. There are opportunities to help editors identify these article webs and find new editing opportunities.

[15] For participants motivated primarily by the research activities involved in creating articles and content, we find Content Translation has less appeal because the tool removes part of that researching experience.

* Parts of visual adapted from 'Slidesgo'

"

Reaching readers

"I feel happy about Wikipedia because one page may be viewed by thousands of people. I recently wrote about Covid-19 and the article was read by 4,000 people" [CE15_KN]

Receiving recognition

"I don't want to be anonymous" [CE12_GU]

"When I complete the article, I feel very happy because I will be the main editor of that article/page" [CE13_KN]

Self-improvement

"I get to learn, so that enriches me and that is why I do this work" [CE7_GU]

"Because I am reading and translating, my English is also developing" [CE39_TE]

Powerful social motivators ...and *d*emotivators

"I'm motivated by my friends, who are more experienced editors" [CE31_PA]

"If someone comments negatively on the articles I've written, I won't go to Wikipedia for the next two days" [CE15_KN]

Challenges (General pain points)

1. **Input challenges are widespread and phonetic, probabilistic options are preferred**

   While particularly salient among potential editors, current editors also face input challenges and have developed a range of different solutions and workarounds. In general, WP keyboard options are not easy to find for newcomers. Overall, the ability to input text phonetically and receive predictive text options is the method participants were most familiar and comfortable with.

2. **Access to online sources for references is very limited**

   Many editors have to go offline to find sources (e.g., CE7_GU, CE39_TE, CE8_GU, CE25_MR). It's very common for editors to need to go to physical libraries to collect resources and spend significant amounts of time digitizing sources. Due to a lack of resources, editors may default to relying on other wikis for sources.

3. **Small wikis present unique editing environments with advantages & disadvantages**

   With fewer editors in general, there's less anonymity and more receptiveness to newcomers who may still be learning norms of editing. At the same time, there are fewer editors to complete admin tasks and proofread articles. This causes delays to processes that are very fast in wikis with very large editor bases.

   Small wiki editors from different regions also face challenges when editing other wikis, such as the English wiki. For example, editors report frustrations that English wiki editors may impose judgements on notability for highly notable topics in other regions. Other big wiki editors may also not be well-positioned to judge more local sources, all of which creates headaches for small wiki editors, whose work may be undone.

4. **Editors face an assortment of technical and conceptual challenges**

   Editors report an assortment of other challenges they faced when beginning to edit, such as:

   - Finding time to learn wiki markup language (e.g., CE14_KN) and learn to edit (e.g., 6-7 months CE7_GU)
   - Being afraid to make mistakes and be judged (e.g., CE31_PA)
   - Adding images from commons and navigating an "old" interface (e.g., CE21_ML, CE39_TE)
   - Plus, problems faced due to internet and device access (including scanners for digitizing sources)

Input challenges are common and *not* CX-specific

"I don't like to type" [PE24_ML]

"How do I get the Punjabi keyboard?" [PE35_PA]

"Yeah, that's my weapon - Google Input tools" [PE18_KN]

"Since my Bengali typing is slow, I will type in English in Google Translate and then get the work done in the Bengali wiki" [CE1_BN]

"I prefer to write content using Google Input Tools. It converts it for me and gives me suggestions related to what I'm expecting" [PE42_TE]

Small wiki, less anonymity

"I was editing the English Wikipedia, but it was too much stress. [...] The Gujarati Wikipedia has two admins and both are friendly. Both welcome us and help if we are stuck" [CE9_GU]

An assortment of technical & conceptual challenges

"The main problem I faced first was the technical rules to be followed on Wikipedia" [CE7_GU]

Where to get help?

Potential editors have little experience navigating Wikipedia

Readers tend to navigate Wikipedia almost exclusively through series of clicks on search result pages, so the landscape beyond article contents is unfamiliar. (e.g., PE12_GU, PE16_KN, PE28_MR, PE29_MR, PE34_PA, PE23_ML)

Newcomers are looking for contextual help

In most cases, research participants faced problems related to a specific task they were trying to complete. They wanted to troubleshoot a very specific, contextual problem; not spend time reading long passages of help contents.

Many newcomers are struggling to even know where to start

Exemplified well by one participant's experience with having his images removed, he didn't know who to contact or who to ask to learn why they kept getting removed (CE1_BN). Put very simply, new editors 'don't' know where to start' in their quests for help.

OPPORTUNITIES

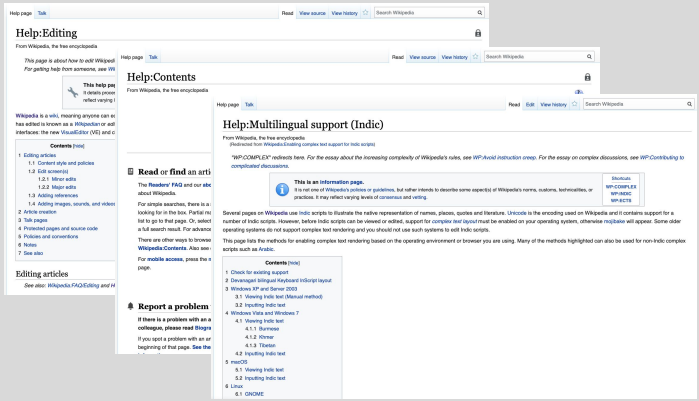
1. **Prioritize contextual help options** in interfaces and workflows so editors don't have to resort to digging through long detailed help pages to troubleshoot very specific problems.
2. **Surface opportunities for editors to connect with support networks.** From understanding there's a network of active editors on their wiki to making connections with other editors using translation, new editors and translators would benefit from social support systems. Via current editors we know these interactions can also help provide motivation to keep editing and work through challenges. Successful, experienced translators on Wikipedia report having networks they tap into to get help with challenging vocabulary. Newcomers don't have these same supports.
3. **Assume help will be needed.** Anticipate newcomer questions when designing new translation and editing interfaces, and scour them for jargon and concepts only advanced editors will understand.

The *'where do I start?'* problem of getting help

"So far, from all the pages I saw, I didn't see anything that could help me" [PE10_GU]

"Something should be mentioned about how to do it, it becomes very complicated when you check the help section" [PE36_PA]

"I've read Wikipedia for years, but never been to their homepage" [PE12_GU]



(Assorted help pages that participants never uncovered in sessions)

Additional Device Patterns

Mobile-first and laptop-first editors

Participants included both mobile-first users and editors who work primarily from laptops and chromebooks. For example, CE2_BN makes over 90% of their edits from a smartphone, whereas CE33_PA feels unable to edit from their phone. Yet, most mobile-first users report difficulty editing on mobile, especially when it comes to longer edits and content such as infoboxes.

Editing and reading activities correlates with device use

Overall, editors reported relying on mobile for 'minor' edits, such as corrections, inserting links, or when traveling and away from their other device(s) (e.g., CE9_GU, CE26_MR, CE38_TE, CE20_ML). Many participants also report preferring to read Wikipedia from smartphones. However, for many editors, devices with larger screen sizes remain preferred for drafting longer content and working with more complex content types, such as infoboxes and formulas. For example, Malayalam editor and participant 20 said, "I mainly use laptop for creations - I don't know how to create new content on mobile."

Input challenges may be less common on mobile devices

A number of participants - especially potential editors - were more comfortable with non-Latin script inputs on their smartphones (e.g., PE18_KN, PE36_PA, PE41_TE). This is because participants were more comfortable adjusting input settings on their smartphones. In the words of Punjabi editor and participant 36, "Keyboards and language changes...all these happen easily on the phone." Conversely, it was not uncommon for the session moderator to need to assist with such adjustments on participants' laptops.

OPPORTUNITIES

1. **Surface translation entry points** with mobile readers in mind because editors contributing with translation often begin by reading on larger wikis and identifying content gaps. Moreover, cater to device switching patterns by allowing editors to save articles to translation dashboards for later work.

2. **Prioritize mobile-first responsive design** to support both types of editors. This also helps set up newcomers for success as it provides easier entry points for new editors who face fewer input challenges on their mobile devices.

3. **Translation suggestions should be tailored with device in mind.** Shorter articles and missing article sections lend themselves more easily to mobile translation. Longer articles and those with lots of complex content types present formidable challenges for editing on mobile. By tailoring translation suggestions, we can better set up editors for success and accommodate their preferences.

Brand awareness doesn't imply knowledge of how editing works
(the 'tip of the iceberg' problem)

For the average reader, the editing process remains below the tip of the iceberg

Even potential editors with years of reading experience are not always clear on who edits Wikipedia and how accessible editing is. A surprising number of potential editor participants in this study had relatively limited knowledge of how editing works on Wikipedia. Upon their first attempts at editing, we frequently observed a reaction of surprise at how accessible they perceived it to be (unfortunately 'accessible' didn't imply 'easy' in most cases). As for specific methods of contribution, all potential editor participants in this study were completely unaware of editing options involving translation tools available on Wikipedia.

Large wikis frequently remain the default reading option for many multilingual readers

Due in part to search results and in part to awareness of how larger wiki contents are more complete, multilingual readers often default to reading larger wikis, such as the English version (e.g., PE12_GU, P23_ML, PE36_PA, PE40_TE, PE24_ML, PE28_MR). In the words of Marathi speaker and participant 28, "The maximum content on Wikipedia is available in English." Participants reported relying on small wikis for content specific to regions.; for example, reading about a Gujarati author or actor. These patterns indicate that for newcomers, there are more chances to surface editing and translation entry points on the big wikis, leading them to editing opportunities on the smaller wikis.

New editors are not immune to social and community effects on satisfaction and motivation

Even after making the conversion from reader to editor, social factors greatly affect their editing behaviors. For example, one participant had stopped using Content Translation due to objections from the community about the way they were translating and publishing. This participant did not understand the specific objections given. Another editor talked about how they self-selected their editing behavior based on anticipated discussions they may need to have with other editors. In some cases, these discussions can be argumentative and not always conclusive.

Despite 100% brand awareness, editing awareness is fuzzy

"I have a vague understanding of who creates content on Wikipedia. Anyone can edit, but there are fact checkers and I don't know who they are and where they're from" [PE12_GU]

"According to my understanding, there are research assistants or scholars that write Wikipedia" [PE16_KN]

"I used to feel that a company controls Wikipedia and we can't disturb it" [PE29_MR]

OPPORTUNITIES

1. **Increase entry points for small wiki editors on the larger of the wikis they're reading.** The larger wikis is where they're spending much of their time reading and exploring.

2. **Provide more opportunities for readers and potential editors to learn about translation as a way of contributing.** From highlighting the role that translation plays in helping create articles, to promoting it's advantages for new editors, it currently exists below the tip of the Wikipedia iceberg for most new editors (and even some experienced ones).

5. Personas & Experience Journey Maps

WIKIMEDIA
FOUNDATION

Personas

Small Wiki Multilingual Editors

WIKIMEDIA
FOUNDATION

Overview

From the research session data, the following 5 personas were constructed, and highlight the experiences of **3 existing** and **2 potential** small wiki contributors - all of whom are multilingual and regularly use translation in various ways to contribute to Wikipedia and/or online platforms.

Dharuna *

Current Marathi Editor — 37 years old

- Pune, India
- Marathi, Hindi, English
- Bachelor of science
- Electrical engineer

Niranjan

Potential Malayalam Editor — 44 years old

- Kochi, India
- Malayalam, English
- Master's degree
- Journalist

Suresh

Current Telugu Editor — 27 years old

- Hyderabad, India
- Telugu, English, Hindi, Kannada
- Master's degree
- Small business owner

Manju

Potential Punjabi Editor — 41 years old

- Ludhiana, India
- Punjabi, English, Hindi
- Bachelor degree
- Teacher

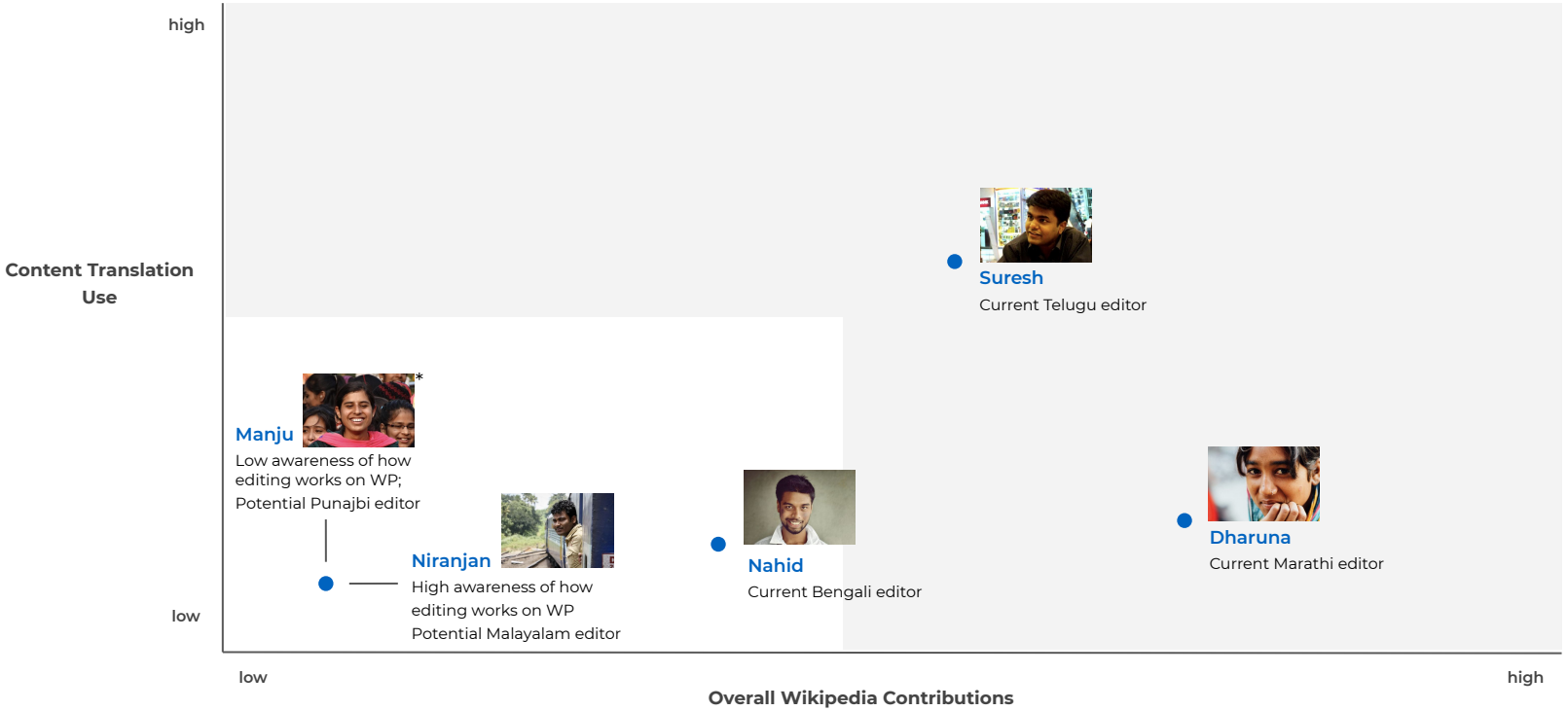
Nahid

Current Bengali Editor — 22 years old

- Kolkata, India
- Bengali, English
- Bachelor degree
- Travel Industry

* Please refer to individual personas for full image credits

Editing & Translation Activity at a Glance



* Please refer to individual personas for full image credits

Dharuna

Current Marathi Editor — 37 years old

Pune, India

Bachelor of science

Marathi, Hindi, English

Electrical engineer

> I try to compare how many people have read my articles and how helpful they've been. If the article I wrote isn't read by anyone, then what's the point?
>
> Women in India have many responsibilities other than just their work - the household, kids, etc. We want to spend our time positively, not engaging in long talk page debates to protect our edits.

Photo by AdamCohn (CC BY_NC_ND 2.0), https://search.creativecommons.org/photos/801af210-1457-43de-adea-7b904c0cc8c9

Editing

Marathi 9,000+ | WIKIDATA 1500 | WIKIPEDIA COMMUNITY 1600 | 1000

Dharuna began editing English Wikipedia in 2014, but switched to editing Marathi Wikipedia after learning of it. Currently, she edits every week, focusing on both producing high quality articles, but also making the smaller edits she feels improves article quality - adding images and references.

She used Content Translation (CX) more as a newer editor after finding out about it in the Village Pump, but less so now. She likes the side-by-side language view of CX, but the machine translation output quality is demotivating. For drafting articles, Dharuna uses a combination of Source Editor, Google Translate, and Google Input Tools. She's accustomed to predictive input methods.

USE OF OUTSIDE TOOLS

few — many

COMFORT WITH MARATHI INPUTS ON WP

low — high

TECH SAVVINESS

low — high

SOCIAL CONNECTIONS IN MARATHI WIKI

few — many

CONTENT TRANSLATION USE

infrequent — frequent

Goals

- Create high quality articles with images and links
- Maintain her reputation as an editor of high quality articles
- Make content available locally for people of her region
- Close the gender gap for articles on Marathi Wikipedia
- Receive recognition for her contributions and efforts

Challenges

- Finding sources
- Translating technical vocabulary
- Editing the Marathi script directly on Wikipedia
- Maintaining language consistency when co-editing articles
- Staying motivated when editing low quality machine translation outputs
- Finding Content Translation
- Protecting her edits from being reverted

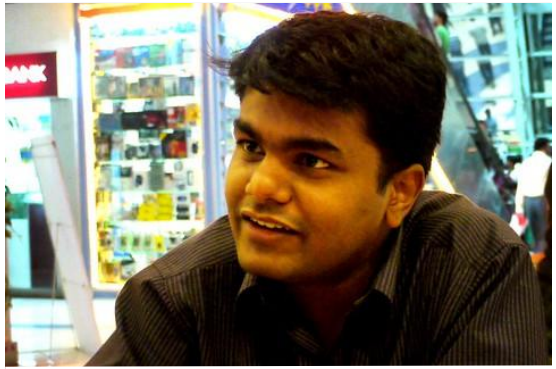
Motivation

- *Wide reach* — knowing people read her articles and connecting with a wide audience through Wikipedia
- *Positive Impact* — doing something positive for her mother tongue and improving coverage of geographically- and culturally-local topics
- *Recognition* — receiving recognition for her editing activity

Suresh

Current Telugu Editor — 27 years old

- Hyderabad, India
- Master's degree
- Telugu, English, Hindi, Kannada
- Small business owner



> By reading English Wikipedia, I can see there are many topics still needed in Telugu Wikipedia. This makes me feel very disappointed ... and then I start writing.
>
> What I enjoy about Content Translation is that it saves my time spent typing in Telugu, and it also adds references.

Photo by legends2k (CC BY 2.0), https://search.creativecommons.org/photos/0b251cd0-7dca-4b7c-8d38-83ee4b265cf1

Editing

Telugu 3,000 · English 250 · 200 · 350

Suresh began editing Telugu Wikipedia in 2017. He identifies articles needed in his wiki by starting at English Wikipedia. His editing activity follows his stream of reading and learning about new topics. He prefers to read on his smartphone, but edits primarily on a laptop. He uses Source Editor for small changes and Visual Editor for editing infoboxes and drafting new content.

Starting at the language selector on an English Wikipedia page, he often uses Content Translation to speed up the process of ensuring Telugu Wikipedia has coverage of current events topics and what he's reading about in local newspapers. If they're short articles, when he feels motivated, Suresh may publish 3-4 articles a day with the tool.

USE OF OUTSIDE TOOLS

few — many

COMFORT WITH TELUGU INPUTS ON WP

low — high

TECH SAVVINESS

low — high

SOCIAL CONNECTIONS IN TELUGU WIKI

few — many

CONTENT TRANSLATION USE

infrequent — frequent

Goals

- Improve Telugu Wikipedia coverage of topics in current events and local newspapers
- Increase the number of articles available in his wiki
- Build social connections and learn to edit more efficiently

Challenges

- Writing in Telugu, especially when editing machine translation outputs
- Editing infoboxes
- Machine translation limits, which block publication of content he feels is good quality
- Working on an article with Content Translation little-by-little when other editors may not know he's working on it and start again
- Images he adds keep getting deleted; he doesn't know why

Motivation

- *Wide reach* — reaching a wide audience through Wikipedia
- *Helping others* — providing access for those who don't read English or Hindi
- *Self-improvement* — learning and improving his English while translating articles
- *Positive feedback from community* — encouraging messages from other editors; knowing he's editing correctly

Nahid

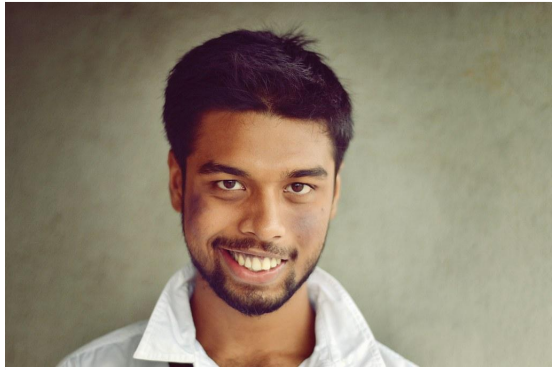
Current Bengali Editor — 22 years old

Kolkata, India

Bachelor degree

Bengali, English

Travel Industry



> What I enjoy most is when others read what I've edited...I feel good.
>
> Work that takes 5 minutes on a laptop takes me 10-15 minutes on smartphone.
>
> I feel that if I continue editing, my talent and skills will increase from it.

Photo by Soumyadeep Paul (CC BY 2.0), https://search.creativecommons.org/photos/312fb012-4da8-42c8-b683-116908901112

Editing

Bengali 450 English 150

Nahid began editing Wikipedia during the Covid-19 pandemic in 2020. He uses a smartphone for about 90% of his editing, and generally focuses on short edits and editing 2-3 paragraphs at a time. He enjoys editing topics that are geographically- and culturally-relevant to where he lives. His inherent knowledge of these topics provides confidence when faced with other editing challenges.

Using Google Translate, he may add 2-3 paragraphs to articles at a time. Typing in Bengali is easy on mobile, but Nahid relies on the Google Translate interface for writing on a laptop. He found out about Content Translation through an editing competition. But, he doesn't know it's out of beta in his wiki and cannot locate the Content Translation dashboard.

USE OF OUTSIDE TOOLS

few — many

COMFORT WITH BENGALI INPUTS ON WP

low — high

TECH SAVVINESS

low — high

SOCIAL CONNECTIONS IN BENGALI WIKI

few — many

CONTENT TRANSLATION USE

infrequent — frequent

Goals

- Create content that readers find valuable
- Increase his skills and make some new friends
- Provide information for people in remote areas who don't speak English
- Keep the Bengali Wikipedia as up-to-date as larger wikis

Challenges

- Writing in Bengali on a laptop (enabling keyboard options)
- Editing infoboxes, especially on a smartphone
- Templates not available in Bengali Wikipedia
- Finding the best words when translating
- Editing more than a short paragraph on his smartphone
- Finding Content Translation

Motivation

- *Self-improvement* — feeling like he's learning new skills
- *Reach* — seeing content he's helped edit appear on Google searches and getting accessed by others
- *Article quality* — knowing an article on Bengali Wikipedia is just as up-to-date as the same article on a bigger wiki

Manju

Potential Punjabi Editor — 41 years old

Ludhiana, India

Bachelor degree

Punjabi, English, Hindi

Teacher

> When I start editing, there should be some help or recommendations, like knowing how to change settings so I can write in Punjabi more easily.
>
> After publishing, I feel good that I wrote something and people will get to see that.
>
> Content Translation turned out to be helpful, I think it's very beneficial when just getting started.

Photo by Adam Jones (CC BY-SA 2.0), https://search.creativecommons.org/photos/ed61de8c-7093-4157-8e53-7e618e1bfab0

Reading & Editing

Manju sees Wikipedia as the place to go get authentic and complete information. She reads both Punjabi and English Wikipedia. For example, she reads English Wikipedia to access information about specific medicines, but prefers reading about general health-related articles on Punjabi Wikipedia. While she understands that many writers are involved in creating Wikipedia, she doesn't understand who creates content and how.

After being invited to try editing, she first thought about editing when reading about a local village, and enjoys correct mistakes - fixing factual errors. Manju has access to a laptop, but generally writes and translates for her personal blog on her phone. One reason for this is because she feels she types fluently in Punjabi on her smartphone, using predictive text options.

When invited to try Content Translation, she had difficulty with keyboard settings and writing. She also faced challenges knowing how to interact with the tool and how to engage with the very long infobox that appeared at the top of the article.

KNOWLEDGE OF HOW WP EDITING WORKS

basic — advanced

COMFORT WRITING IN LOCAL LANGUAGE(S)

low — high

TECH SAVVINESS

low — high

Goals

- Create visually-rich content
- Receive recognition for editing and writing
- Benefit large groups of readers
- Be part of a well-known and respected platform
- Understand her impact (feedback loop)
- Share her contributions with friends

Challenges

- How to start editing Wikipedia?
- Understanding who creates content
- Getting help understanding the process: how is content reviewed, where does she draft and save?
- Feeling confident about her content knowledge
- Adopting what the community would consider a neutral stance
- Feeling anonymous

Motivation

- *Positive feedback & recognition* — getting feedback from readers and recognition for her efforts
- *Wide reach* — a feeling of satisfaction from reaching a wide audience on Wikipedia
- *Writing to learn* — suddenly watching something that makes her want to learn and write about it

Niranjan

Potential Malayalam Editor — 44 years old

Kochi, India

Master's degree

Malayalam, English

Journalist

"

If it's not user-friendly, I won't use it.

Content Translation provides formatting, photos, and links. You can't get that from Google Translate.

Even a 50-year-old man can use Facebook easily. I expected that Wikipedia would be as easy, but it's not. It takes much more effort to learn.

Photo by Fabionik (CC BY-NC-SA 2.0), https://search.creativecommons.org/photos/41ff6286-84da-4655-9a90-24bce5dc2a45

Reading & Editing

Niranjan generally defaults to reading English Wikipedia, except for learning about local authors, films, and regionally-famous people. He rarely navigates within Wikipedia and instead relies on Google searches to navigate the site.

He has a good understanding of how anyone can contribute to Wikipedia, but a more limited technical understanding. Niranjan uses a Google search to find an English article he can translate to Malayalam. His first editing experience involved composing in Google Input Tools and pasting into Visual Editor. Outside of Wikipedia, he's accustomed to the interfaces of WordPress, Quora, and Medium.

Only with assistance did Niranjan find Content Translation (CX). He likes how it can translate simple sentences very quickly, and how links and references are automatically inserted. Two major barriers to use are feeling like it was too much work to edit machine translation output, and knowing 'what to do next'.

KNOWLEDGE OF HOW WP EDITING WORKS

basic — advanced

COMFORT WRITING IN LOCAL LANGUAGE(S)

low — high

TECH SAVVINESS

low — high

Goals

- Contribute longer form articles
- Be able to draft articles in a way that feels familiar to current writing processes
- Receive recognition or credit for contributions
- Understand who co-contributors are and the 'behind-the-scenes' of articles

Challenges

- Adding citations
- Saving unfinished drafts to complete later
- Navigating the Wikipedia ecosystem
- A slow connection, causing tools like Content Translation to load very slowly
- Writing in Malayalam without predictive phonetic input tools

Motivation

- *Wide reach* — reaching a wide audience is very satisfying
- *Prestige* — participating as a writer on a platform seen by thousands
- *Feedback* — receiving feedback from readers is highly motivating

Personas — Definition of Quick Matrices

Current editors

USE OF OUTSIDE TOOLS

Use of Wikimedia-external tools while writing and editing

COMFORT WITH LANGUAGE INPUTS ON WIKIPEDIA (WP)

Degree of exclusive reliance on WP input tools vs. use of external tools to aid input

CONTENT TRANSLATION USE

Frequency of Content Translation use for writing articles

SOCIAL CONNECTIONS IN SMALL WIKI

General activity in social areas, including talk pages, Whatsapp or Facebook groups, and in-person gatherings

Potential editors

TECH SAVVINESS

General level of tech savviness on/off Wikimedia projects

KNOWLEDGE OF HOW EDITING WIKIPEDIA (WP) WORKS

Awareness of how content is created and ability for newcomers to edit

COMFORT WRITING IN LOCAL LANGUAGES

General level of comfort writing in non-Latin scripts, especially for most relevant local language(s)

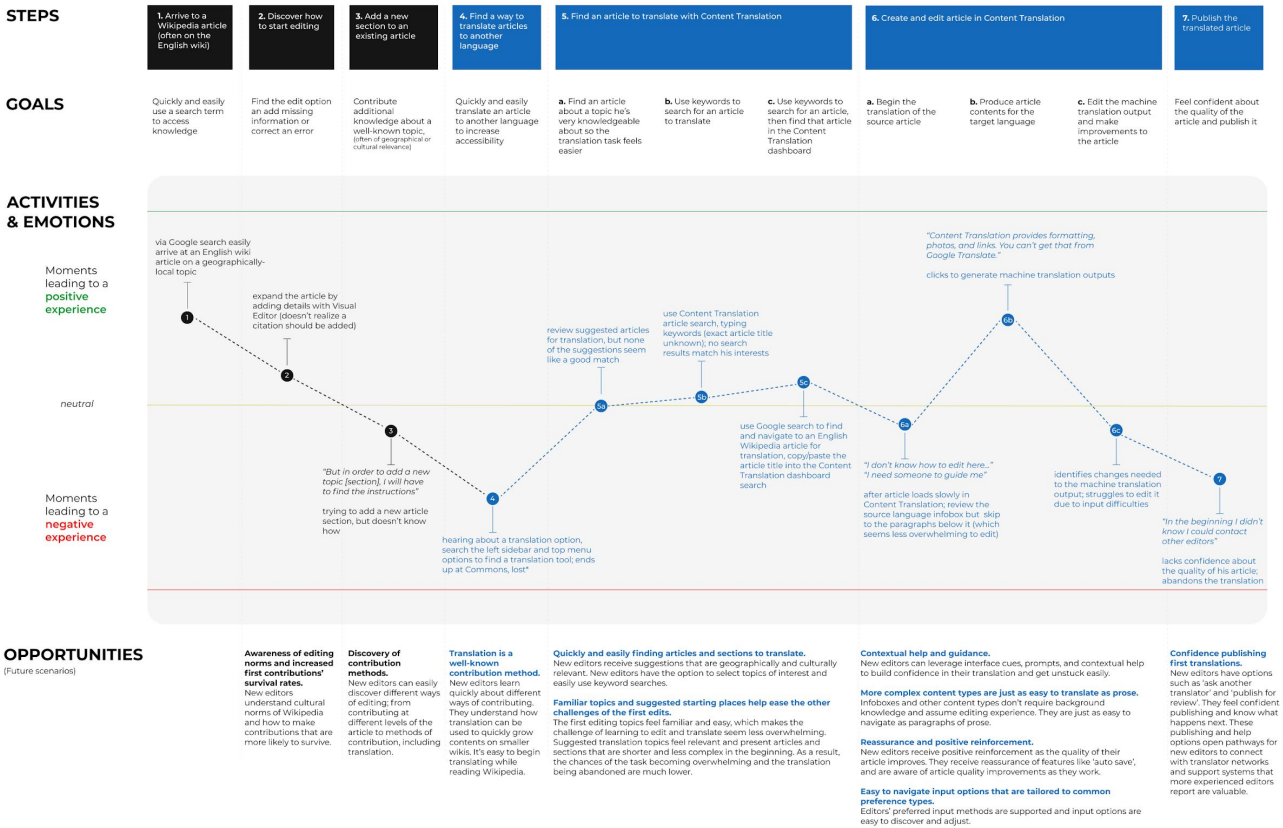
Experience Journey Maps

Small Wiki Multilingual Editors

WIKIMEDIA
FOUNDATION

New Translator Journey Map

Niranjan is a first-time Wikipedia editor who speaks Malayalam and English. He generally accesses the Malayalam Wikipedia to learn about local authors, films, and regionally-famous people, which is where his editing journey begins. After hearing about an option to edit with translation, he struggles to discover how to do this. Eventually, he decides to translate an English article. The high point of his experience was discovering all the editing tasks Content Translation does automatically. The lowpoints of his experience were knowing where to begin and facing difficulty editing the machine translation output, the latter of which contributed to him not publishing his article.



OPPORTUNITIES

(Future scenarios)

Awareness of editing norms and increased first contributions' survival rates.
New editors understand cultural norms of Wikipedia and how to make contributions that are more likely to survive.

Discovery of contribution methods.
New editors can easily discover different ways of editing; from contributing at different levels of the article to methods of contribution, including translation.

Translation is a well-known contribution method.
New editors learn quickly about different ways of contributing. They understand how translation can be used to quickly grow contents on smaller wikis. It's easy to begin translating while reading Wikipedia.

Quickly and easily finding articles and sections to translate.
New editors receive suggestions that are geographically and culturally relevant. New editors have the option to select topics of interest and easily use keyword searches.

Familiar topics and suggested starting places help ease the other challenges of the first edits.
The first editing topics feel familiar and easy, which makes the challenge of learning to edit and translate seem less overwhelming. Suggested translation topics feel relevant and present articles and sections that are shorter and less complex in the beginning. As a result, the chances of the task becoming overwhelming and the translation being abandoned are much lower.

Contextual help and guidance.
New editors can leverage interface cues, prompts, and contextual help to build confidence in their translation and get unstuck easily.

More complex content types are just as easy to translate as prose.
Infoboxes and other content types don't require background knowledge and assume editing experience. They are just as easy to navigate as paragraphs of prose.

Reassurance and positive reinforcement.
New editors receive positive reinforcement as the quality of their article improves. They receive reassurance of features like 'auto save', and are aware of article quality improvements as they work.

Easy to navigate input options that are tailored to common preference types.
Editors' preferred input methods are supported and input options are easy to discover and adjust.

Confidence publishing first translations.
New editors have options such as 'ask another translator' and 'publish for review'. They feel confident publishing and know what happens next. These publishing and help options open pathways for new editors to connect with translator networks and support systems that more experienced editors report are valuable.

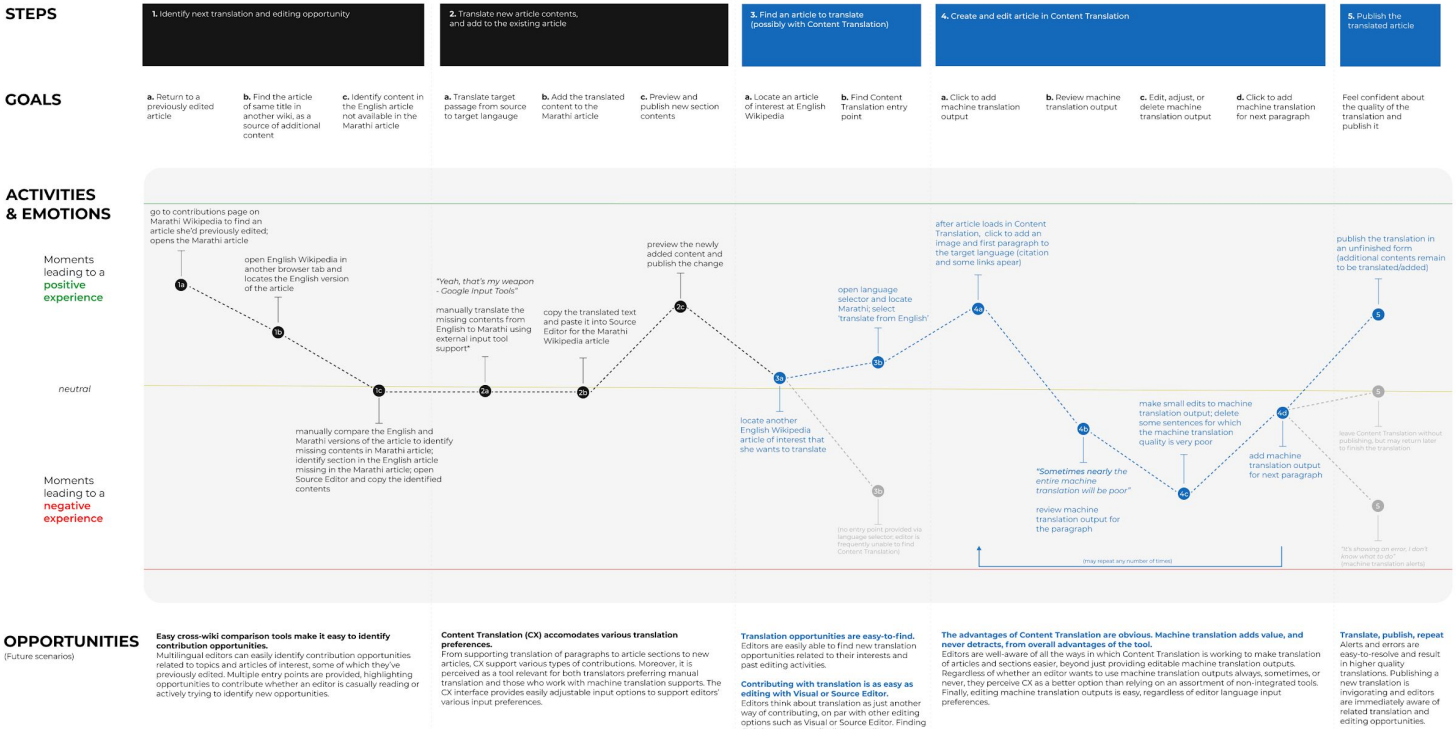
* Requires direct assistance from someone in order to find how to access Content Translation

Access this file on Commons

https://commons.wikimedia.org/wiki/File:New_Translator_Journey_Map.png

Experienced Translator Journey Map

Dharuna began editing Marathi Wikipedia shortly after starting to edit English Wikipedia in 2014. She focuses on producing high quality articles with images. She used Content Translation more frequently as a newer editor, but has since come to rely on some of her own translation workflows. Translation is a regular editing activity, and her process typically involves multiple browser tabs, source editor, and Google Input Tools. Although she doesn't always translate with Content Translation, she really enjoys the side-by-side source-target language layout and that citations get added automatically. The main drawback for her is editing the machine translation output.



OPPORTUNITIES
(Future scenarios)

Easy cross-wiki comparison tools make it easy to identify contribution opportunities.
Multilingual editors can easily identify contribution opportunities related to topics and articles of interest, some of which they've previously edited. Multiple entry points are provided, highlighting opportunities to contribute whether an editor is casually reading or actively trying to identify new opportunities.

Content Translation (CX) accomodates various translation preferences.
From supporting translation of paragraphs to article sections to new articles, CX support various types of contributions. Moreover, it is perceived as a tool relevant for both translators preferring manual translation and those who work with machine translation supports. The CX interface provides easily adjustable input options to support editors' various input preferences.

Translation opportunities are easy-to-find.
Editors are easily able to find new translation opportunities related to their interests and past editing activities.

Contributing with translation is as easy as editing with Visual or Source Editor.
Editors think about translation as just another way of contributing, on par with other editing options such as Visual or Source Editor. Finding CX is just as easy as finding the edit article/section option.

The advantages of Content Translation are obvious. Machine translation adds value, and never detracts, from overall advantages of the tool.
Editors are well-aware of all the ways in which Content Translation is working to make translation of articles and sections easier, beyond just providing editable machine translation outputs. Regardless of whether an editor wants to use machine translation outputs always, sometimes, or never, they perceive CX as a better option than relying on an assortment of non-integrated tools. Finally, editing machine translation outputs is easy, regardless of editor language input preferences.

Translate, publish, repeat
Alerts and errors are easy-to-resolve and result in higher quality translations. Publishing a new translation is invigorating and editors are immediately aware of related translation and editing opportunities.

*Overall, existing translators were observed using various external tools for manual translation and drafting content. Other tools observed for this purpose included MSWord, Google Translate (but primarily for input, not translation), and other 3rd party input assistant websites.

Access this file on Commons
https://commons.wikimedia.org/wiki/File:Experienced_Translator_Journey_Map.png

6. Wrapping up & looking ahead

Key takeaways

Next steps

WIKIMEDIA FOUNDATION

Executive Summary - *Top 10 Key Takeaways*

1. Most participants, including both current and potential editors, were unable to successfully navigate to Content Translation (CX), despite consistently searching in many of the same places. Many pre-translation workflows begin at large wikis where editors are reading.
2. Current editors and newcomers are delighted by a number of CX features that speed up the editing process and require less typing, a challenge for some users. Delighters included automatic links and references, side-by-side presentation of the source/target article, and help generating key article vocabulary.
3. Keyboard input challenges with non-Latin scripts are widespread. This is not a CX-specific problem, but impacts the CX experience in critical ways. While some current editors have mastered non-Latin inputs on WP, others have developed a range of workaround solutions. Without the ability to type fluently, users are unable to edit machine translation outputs.
4. Observation of editing workflows revealed that 17 of 21 current editors sampled relied on external tools for at least one of the following purposes: input, drafting, translation. These external supports are helping editors accomplish goals such as entering non-Latin scripts, drafting new articles, and translating article sections.
5. Alerts/warnings in Content Translation are not always well-understood and timing of their presentation is critical. Editors didn't always understand how to to resolve alerts, and are not clear on how to proceed if ready to publish but not able to due to a machine translation use limits.
6. An editor's motivation is multi-dimensional. Motivations range from purely altruistic in nature to those grounded in self-improvement. During research sessions, we observed 6 primary patterns: improving access, expanding contents, self-improvement, social, audience, recognition/identity. (more details, slide 24)
7. Small wikis present unique environments for editors compared to larger wikis. For example, editors noted that when compared to the English wiki, there's often much less anonymity and more receptiveness to new editors. On the other hand there are fewer editors to complete admin tasks, proofread articles, and help find sources.
8. In most cases, participants faced problems related to a specific task. They wanted to troubleshoot a very specific, contextual problem, not access a lengthy help page. They're looking for contextual help, and new translators need help building the support systems that experienced translators report having for help and motivation.
9. Small wiki editors vary in terms of whether they are primarily mobile-first editors or rely on laptops/chromebooks. In general, editing and reading activities correlate with device use patterns. In general, we found that input of non-Latin scripts is slightly easier for potential editors on mobile devices.
10. Despite nearly 100% brand awareness, the editing process remains below the tip of the Wikipedia iceberg. Multilingual readers often default to reading on larger wikis, and translation as a form of editing lacks widespread awareness - even among current editors. Social factors are in part determining how editors contribute and what topics they contribute about.

Next steps

Ideation, Design, Evaluation

Ideation, Prioritization, & Planning

Using the research deliverables from this project, the Language Team can focus on ideation, prioritization, and planning for work in upcoming quarters and fiscal years. The findings around CX discoverability motivate current efforts underway to prioritize improving entry points and discoverability.

Scoping & cross-team collaboration

Some of the outcomes of this work point to work that will fall both within and outside the scope of the language team. Thus, there are opportunities to coordinate and/or communicate with other teams on issues impacting small wiki editors and translation.

Evaluation of Section Translation (SX)

As SX becomes available in wikis in 2021, evaluate how SX is meeting the needs of mobile-first users, and compare its adoption among different groups of users. Combine analytics with user research to identify gaps and opportunities for SX users across all devices.

Research

Section translation (SX) early adopter feedback

Usability testing of Section Translation and feedback from early adopters. Currently in a test server and soon to be released into its first wikis, SX aims to solve some of the current limitations of Content Translation. It enables mobile translation and translation of article sections, opening up opportunities for more collaboration on the translation of articles.

Content translation (CX) entry points research

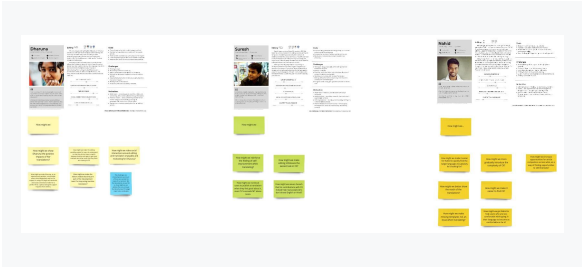
The current project underscores the need for a greater number of easily discoverable entry points for Content Translation. Research should support design exploration and prioritization of entry points, helping to determine which are most promising.

Machine translation, human editors

As the number of translators increases, it will be important to continue ensuring high quality articles and sections are published. An initial proposal for this research lays out plans to understand how users edit MT outputs with the goal of improving the MT limits system of alerts and how they're displayed to users.

Power CX user profiles

A case study will help identify what factors contribute to the success and productivity of Wikipedia's most active CX users?



Thanks to:

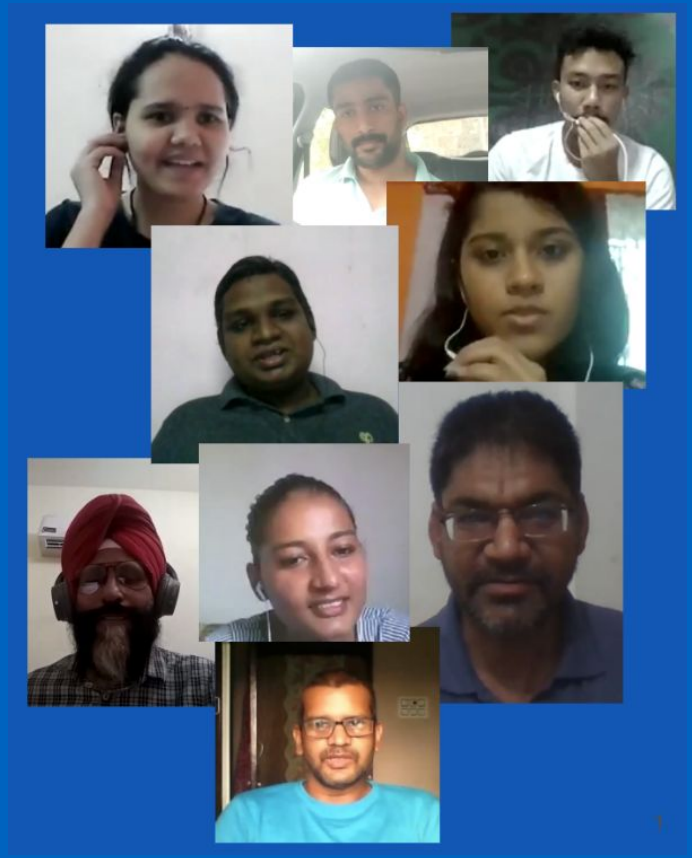
Research Participants - for their time, honest feedback, and inspiring work helping smaller wikis grow

Wikimedia Foundation Language Team - ongoing feedback & discussions

Bethany Gerdemann - research support

Design Strategy - research support, feedback, and discussions

Anagram Research - supplementary recruitment efforts and non-English session moderation



WIKIMEDIA
FOUNDATION

Thank you

Product | Design Strategy | Language Team

Please direct questions & comments to:
eli@wikimedia.org

WIKIMEDIA
FOUNDATION

Appendix

Full rationale for wiki selection

WIKIMEDIA
FOUNDATION

Full rationale for wiki selection (1 of 3)

As part of their annual plan, the Language Team used a number of criteria to identify a set of small wikis to focus on - small wikis with significant potential for growth. 'Small' was defined as wikis with less than 100k articles and less than 100 translations per month. The growth potential can be measured by looking at both active editing behavior and overall size of the language community. More specifically, there are small wikis with more than 70 editors (and more than 20 active editors), and many of these wikis are in languages with millions of speakers. In other words, while not every speaker represents a potential editor, significant segments of the world's population nonetheless that would benefit from content creation and knowledge sharing.

The initial set of target 'boost' wikis, as proposed in the Language Team's annual plan, is shown to the right.[a]

Numbers for these languages were updated on 27 March 2020, to provide more equal comparison with the data for 'boost-compatible' languages, also presented in this report on the next slide.[b] Approximately 70% of active editors across this set of 6 core boost initiative wikis are comprised of Malayalam and Bengali editors; the remaining 30% are represented by the other 4 wikis. As for total speaker bases, 75% of speakers are comprised of Malayalam and Bengali speakers.

Core Boost Initiative Communities (n=6)

ID	Name	Trans.	Articles	Users	Active editors/mo	MT	Mob. views	Speakers
ML	Malayalam	4.9k	68K	135k	90 (+15/mo)	✓	76%	38M
BN	Bengali	7.6k	84k	273k	266 (+75/mo)	✓	86%	265M
TL	Tagalog	2.5k	72k	105k	34 (+4/mo)	✓	66%	25M
JV	Javanese	1.7k	57k	43k	41 (+13/mo)	✓	37%	68M
MN	Mongolian	157	19k	64k	22 (+5/mo)	✓	50%	3.6M
SQ	Albanian	2.9k	79k	122k	59 (+12/mo)	✓	69%	7.5M

[a] https://www.mediawiki.org/wiki/Content_translation/Boost

[b] Numbers were updated by retrieving data from the following sources: Number of articles and users come from data at https://meta.wikimedia.org/wiki/List_of_Wikipedias; number of translations can be found at https://en.wikipedia.org/wiki/Special:ContentTranslationStats; data on mobile (web) (page)views and monthly active editors (monthly new active editors) is found at https://docs.google.com/spreadsheets/d/1a-UBgsYtJl6gpauJyanx0nyxuPqRvhzJRN817XpkuS8/edit#gid=1861872875; finally, total speakers were retrieved from Ethnologue when not behind a paywall https://www.ethnologue.com/guides/ethnologue200; Ethnologue figures represent combined first and second language speakers; all other speaker population data was taken from the languages' English Wikipedia pages.

Full rationale for wiki selection (2 of 3)

In addition to the core set of boost languages, the team also identified a peripheral set of 'boost compatible' languages.[c]

In light of the 2020- pandemic, to maximize flexibility and ensure success for the project, in selecting the final set of focus wikis for the project, we considered both 'core boost' and 'boost compatible' languages as a starting point.[d]

Thus, the project focused on a subset of these boost initiative compatible languages presented in the chart to the right and the core boost initiative wikis on the previous slide. This final subset is presented on the following slide with some additional notes.

Boost Initiative Compatible Languages (n=12)

ID	Name	Trans.	Articles	Users	Active editors/mo	MT	Mob. views	Speakers
TE	Telugu	1.5k	70k	95k	49 (+10/mo)	✓	80%	93M
AF	Afrikaans	2.2k	90k	121k	40 (+2/mo)	✓	43%	18M
KN	Kannada	1.4k	26k	62k	37 (+4/mo)	✓	80%	56M
IS	Icelandic	135	49k	72k	30 (+5/mo)	✓	29%	314k
LV	Latvian	134	101k	90k	89 (+11/mo)	✓	44%	1.8M
BS	Bosnian	101	82k	123k	44 (+7/mo)	✓	56%	3M
GU	Gujarati	47	29k	56k	16 (+3/mo)	✓	75%	61M
NE	Nepali	43	34k	47k	29 (+3/mo)	✓	44%	25M
MR	Marathi	385	57k	120k	69 (+23/mo)	✓	87%	95M
MK	Macedonian	236	105k	87k	60 (+8/mo)	✓	48%	2.5M
SW	Swahili	36	58k	41k	26 (+5/mo)	✓	72%	99M
PA	Punjabi	8.3k	34k	32k	35 (+6/mo)	✓	50%	116M

[c] For as equal comparison as possible, the data for the 'boost compatible' languages was retrieved in the same way as for the 'core boost' wikis, and on the same date (described in footnote [b].

[d] The pandemic introduced a great number of unknown factors going into the data collection phase, such as how available and willing potential research participants would be. All sessions also had to be performed in-home, which introduced challenges for any participants who normally leverage a form of public internet access, such as a library, cafe, or university. These are just a few examples we identified at the onset when planning.

Full rationale for wiki selection (3 of 3)

Finally, to arrive at the final set of wikis selected for the project, we focused on a subset of the wikis presented on the previous 2 slides.

We considered metrics such as speaker population size, number of active editors, number of translations, and general feasibility and research cost effectiveness.

Speaker population size was important because it is a partial rough proxy for potential growth. While not every speaker of a language represents a potential Wikipedia contributor and translator, when examining the boost and boost compatible languages, 7 of the top 10 by speaker size are spoken in India (a cumulative 724M). In addition to large speaker population size, 4 of the top 10 in terms of most active monthly editors are also languages spoken in India. This not only represents existing activity, but also ensures that on the research front we would be able to recruit existing editors from these small wikis. Combine all of this with the fact that 6 of the top 10 in terms of total number of translations are also languages spoken in India, and we arrive at a set of motivations for focusing research efforts on a set of boost and boost compatible languages spoken in India.*

Final selection of wikis for current project

Language	Translations	Articles	Users	Active editors/mo	MT	Mobile views	Speakers
Telugu	1.5k	70k	95k	49 (+10/mo)	✓	80%	93M
Malayalam	7.6k	68k	135k	90 (+15k/mo)	✓	76%	38M
Kannada	1.4k	26k	62k	37 (+4/mo)	✓	80%	56M
Bengali	7.6k	84k	273k	266 (+75/mo)	✓	86%	265M
Gujarati	47	29k	56k	16 (+3/mo)	✓	75%	61M
Marathi	385	57k	120k	69 (+23/mo)	✓	87%	95M
Punjabi	8.3k	34k	32k	35 (+6/mo)	✓	50%	116M

* The fact that through this process we arrived at the option of focusing research on a single country covering several languages had many advantages. However, we recognize that as a result of this, there may be some country- and/or cultural-specific factors that do not fully generalize to a broader, more global set of small wikis. In future iterations of research,, we aim to expand beyond wikis and languages spoken in India, strategically focusing on additional subsets of the boost and boost compatible languages presented in this report.