Section Translation Usability Testing

Many Wikipedia contributors edit articles via mobile devices. Section Translation brings translation support to mobile device editors, and this project provided usability testing as the tool became available in the first wiki, **Bengali***.*

We observe high mobile usage across many language versions of Wikipedia, including relatively smaller wikis such as Bengali, Malayalam, and Tagalog. Currently, translation of Wikipedia articles is supported for desktop editors via [Content Translation](#), which has been used to create more than 800k articles. [Section Translation](#) brings many of the capabilities of Content Translation to mobile editors, and allows contributors to add individual sections of articles by translating and editing content from another language version of Wikipedia. This project provided usability testing of Section Translation as soon as it was available for Bengali editors to expand articles in their language wiki by translating sections from various source language wikis.

Goals   **Results**   **Key Takeaways**

Research Goals

The primary goal of this project was to evaluate the general usability of the Section Translation experience and workflow, in order to inform subsequent product decisions, including prioritization of improvements. In particular, we wanted to know what the main points of friction were that may in part or full prevent editors from making successful contributions. Relatedly, where is better support needed through the process?

The project sought to collect data on the following aspects of the process:

➔ **Selecting** an article and article section to translate

➔ **Reviewing** section contents in the source/target language and initiating the translation

➔ Reviewing, **editing**, and adding to the machine translation output generated by the tool

➔ Completing the translation and **publishing**

Research approach

Methodology

We used moderated remote research sessions to test key aspects of the current Section Translation mobile experience.[1] All sessions were guided by a detailed protocol and involved four main components. Each session began with a short interview. Following, participants shared their screen and used the think-aloud technique while translating an article section on the Bengali Wikipedia. Following observation of their use of Section Translation, participants were asked to complete four Likert scale ratings, and then the session wrapped up with a brief interview

---

[1] Although the experience is in ongoing development and some aspects may change, for reference, it can be accessed on the <u>Bengali Wikipedia here</u>. Because the experience remains in flux, when relevant this report provides screenshots and descriptions of functionality alongside discussion of some results.

period. An interim summary was provided to the Language Team after roughly half of the sessions had been completed.

Participants and data

Participants were recruited from the existing community of Bengali editors through the use of community pump announcements, direct recruitment messages to contributor talk pages, and referrals. Interested participants responded by completing a screener questionnaire, which was followed by an invite to select a session date/time. A total of seven sessions were conducted through the use of a consecutive interpreter, and two were conducted in English.

Most participants resided in Bangladesh at the time of the study, with the exception of one who resided in India and one based in Europe. Seven of the participants reported editing the Bengali Wikipedia on a daily basis; one edited on a weekly basis, and the other had more intermittent editing activity. All participants had experience editing from a mobile phone, and the majority edit multiple times each week from a mobile device. All but two participants regularly use translation to contribute to Wikipedia, with frequency ranging from daily to weekly to monthly. At the time of their sessions, three participants had previously tried Section Translation, whereas six were experiencing it for the first time. All participants joined the session and used Section Translation from a mobile phone.

All sessions except one were recorded (audio and screenshare), and the process resulted in a total of approximately 8.5 hours of video. From these videos (and notes for one participant) we confirmed tracking of key metrics of interest (reported below), and extracted additional data points. These additional data points were coded for thematic analysis. The results section below presents these themes alongside the key metrics we tracked during sessions.

Results

First we'll examine five key topics that reflect key aspects of the Section Translation workflow. Following, a few additional themes are presented, including a review of general feedback.

Topic 1 | Selecting an article

Are editors quickly and easily able to identify and select an article to translate?

The first step to translating a new article section is identifying an article that has missing sections. At the time of testing, aside from manipulating the URL directly, editors were only able to select from the list of suggested articles.

Overall, 7 of 9 participants were not able to find an article from the list of suggestions that was suitable for them. For the purposes of the research session, they were encouraged to select one of these articles anyways so the session could continue, but this is a notable barrier to use of the tool.

So, why didn't they find an article they wanted to translate in the list of suggestions? There were two main reasons. First, the list of suggestions failed to present an article that matched their topics of interest (e.g., (P)articipant5, P7, P8). Secondly, participants noted that they'd already translated articles presented to them as suggestions (P1, P3, P4, P6, P7, P8). P1 confirmed that three suggested sections had already been translated. For some participants, it may not have been clear that there were additional missing sections for the articles they'd previously translated parts of (which were appearing on the list of suggestions). As demonstrated by the quote from P6 below, for any current CX users, some expectations may be driven by their experience with CX. For example, in this example, P6 wondered why they were being suggested article titles that they'd previously worked on.
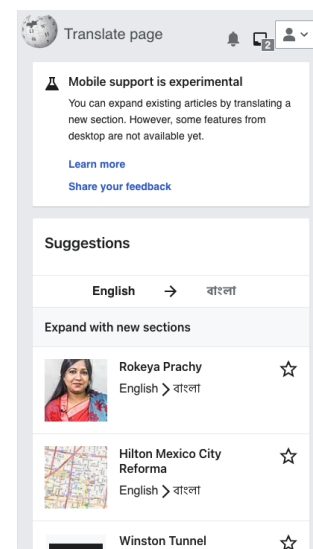
"I don't see anything new that I can translate" - [P4]

"My question is how can I start a new page because the articles they're showing, I already translated them" - [P6]

"I want to translate a particular article because that is relevant to me, but with this [Section Translation] I can't" - [P1][2]

The importance of the quote from P1 above about finding an article "that is relevant for me" captures a key aspect of translation behavior. Overall, participants don't describe translating due to a love of the act of translation; instead, for most it's a means to an end: to learn about a subject or make content available to readers. Translation is driven by enjoyment and pleasure. In the words of P7, "I do whatever I like." For this reason, being able to translate a topic/article of interest is critical to a successful use of Section Translation.

Regardless of whether the list of suggestions failed to present topical interests of the editor or because the editor believed the article had already been fully translated, participants very frequently wanted to search for an article to translate. Moreover, many editors begin their translation workflow directly at the source wiki. For example, P1 begins translation work by looking for cinema-related articles on English Wikipedia. Others, including P4 and P6, began at English Wikipedia, as described by P4 in the quote below.

"...it doesn't give me any option to search for an article I like" - [P1]

"I open an English article and I start translating it to Bengali" - [P4]

In addition to frequently requesting the ability to search for specific articles to work on, another participant (P8) requested that suggestions not be limited to the articles the editor has previously published, as it felt very limiting.

---

[2] Brackets contain information used to add clarification and context to quotes.

After selecting an article, an editor views the screen shown to the right. For at least three of the participants, there was a point of friction observed at this screen (P3, P4, P5). This point of friction was caused by the yellow banner 'already exists' alert. This alert (also used in Content Translation) is intended to signal that the *article* (but not all sections) already exists in the target wiki, but was interpreted as meaning that it already existed in its entirety on the target wiki. All but one participant recovered from their confusion, proceeding by tapping on 'Select section to translation'; however, one participant unnecessarily returned to the list of article suggestions, believing they needed to select a different article.
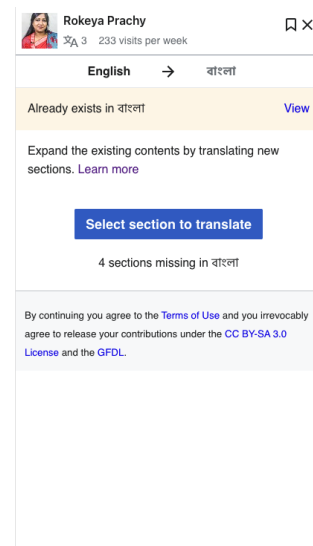
Topic 1 | Key takeaways | Selecting an article

- **Translators want to search for specific articles**
  Without the ability to translate articles of specific interests they perceive as new translation opportunities, editors face a near hard stop with Section Translation.

- **Editors' workflows begin at the source wikis**
  Regarding Section Translation entry points, translators generally begin their work at a source wiki where they are browsing and searching for articles that match topical interests.

- **Current suggestions may be too limiting**
  Although the ability to search for articles/sections to translate should be the foremost priority, there are opportunities to improve how suggestions are generated.

Topic 2 | Selecting an article section and reviewing the contents

Are editors quickly and easily able to identify and select an article to translate? Do they understand the purpose of the step for reviewing the article sections in the source and target language, and does it provide adequate context?

After selecting an article, Section Translation prompts translators to select an article section and on the following screen review the existing source language section alongside the full existing article in the target language (minus the section that the editor wants to add).

Ease of section selection was confirmed through the post-task ratings for the question, *"How would you rate the task of selecting a section to translate?"* Of the participants who responded to this question, 6 of 8 rated section selection as 'easy', while one responded neutrally, and one reported it as slightly difficult.

How would you rate the task of selecting a section to translate?
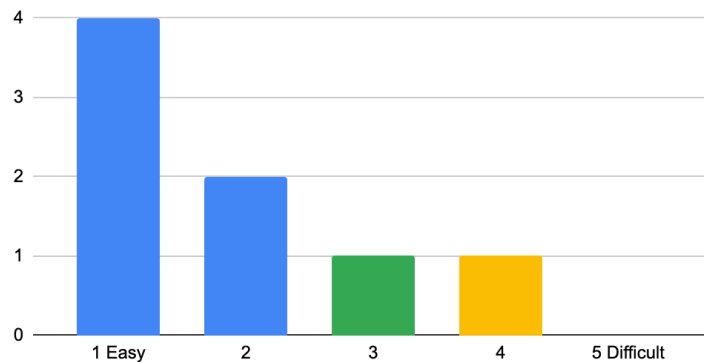
The main barrier to selection tasks, as noted above, was being able to select an article that was relevant to the editor's interests and that they were confident hadn't already been translated. Meanwhile, section selection generally proceeded much more quickly as long as there were 'core'/content sections available to translate. So, if only presented with closing section options such as 'references' or 'external links', participants were not as motivated to begin translating as when they were presented with sections more central to the primary content of the article.

With awareness of the need for more content on their wiki, Bengali editors were most motivated to add sections that provided the best value for anyone trying to

generally understand the topic of the article. All participants were observed preceding linearly, top-to-bottom, in selecting sections to work on.
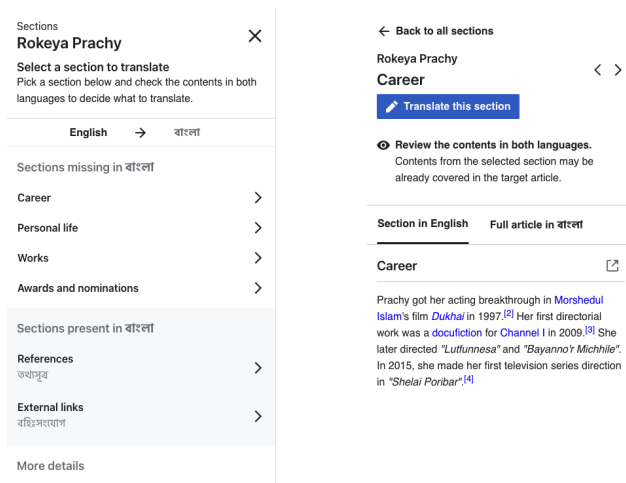
After selecting an article section, editors are prompted to review the selected section in the source language by default, with the option of tapping to preview the full target language article, minus the section they've selected to add.

Approximately half of participants tapped and reviewed the target language article contents (P4, P5, P7, P8). Of these editors, four gave a very brief review before continuing, and one (P5) interpreted the preview of the target language contents to mean that the section already existed. This unnecessarily prompted them to return and select a new article. This behavior was similar to what was occasionally observed for the 'already exists' banner, as noted in the discussion of Topic 1.

Topic 2 | Key takeaways | Selecting a section and reviewing the contents
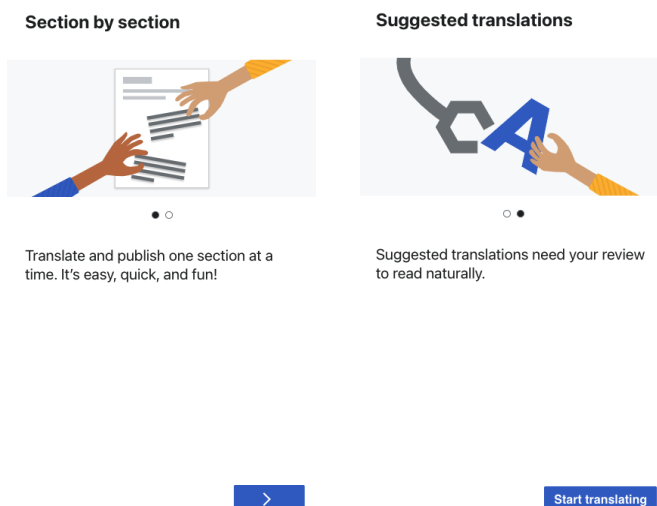
- **Section selection was relatively seamless**
  Overall, section selection presented fewer points of friction when compared to article section.

- **Section selections must present a 'core' content part of the article**
  In order to present a compelling translation opportunity, missing sections must pertain to core content of the article and topic. Editors were less likely to feel compelled to translate closing sections perceived as peripheral, such as 'references' and 'external links'.

- **Content review provides value but the primary 'Translate section' button must remain prominent**
  Because new Section Translation users are still learning the workflow, the 'translate section' button should remain prominent and pinned to the top of the screen to ensure clarity of the next step.

Topic 3 | Quick tutorial

Do the quick tutorial screens provide value and do editors engage with them?

Section by section

Translate and publish one section at a time. It's easy, quick, and fun!

Suggested translations

Suggested translations need your review to read naturally.

Start translating

Overall, participants tapped through these two screens very quickly. At the time of testing these screens were shown universally, for every section translated. The very brief interactions for three of the participants who had previously used Section Translation may be attributed to the fact that they'd seen these tutorial screens before. The only spontaneous question observed at this point in the workflow was from P4 about "how to translate," but their question was immediately resolved when arriving at the next screen, on which the editing task begins. When prompted to recall these screens after tapping through them, P2 described them as "instructions", noting that they expected to get a "raw translation to edit".

Given that participants universally tapped through these screens very quickly with minimal questions, and overall some participants described the Section Translation

process as containing "many screens" or "many steps" (discussed later in this report), it's recommended that some basic logic be implemented to reduce the frequency with which these screens are encountered. They would provide the most value with least chance of annoyance for either a (1) new Section Translation user or (2) Section Translation user who has not used the tool for a significant period of time, potentially multiple months or more. Therefore, one option would be to only display for the first few translations and/or if Section Translation can detect a significant lapse in time since prior use.

Topic 3 | Key takeaways | Quick tutorial

- **Interactions with the quick tutorial were very brief**
  Few if any questions arise for editors at this point in the process. The frequency with which these are shown can be reduced to avoid annoyance.

- **More questions were observed around article and section selection than around what was coming next**
  Thus, focus should be on prioritizing improvements to the article and section selection process to ensure participants get this far in the process.

Topic 4 | Editing translations

Are editors able to understand and successfully leverage all interactions needed to review and edit machine translation outputs?

All but one participant successfully navigated through the editing process, and 8 of 9 participants understood that the workflow proceeded sentence-by-sentence.

Moreover, all participants edited the MT outputs, but none were observed adding additional contents to the article outside of the source article contents. Overall, as shown below, all but one participant rated the process of editing the machine translations as easy.

How would you rate your experience editing the machine translation outputs?
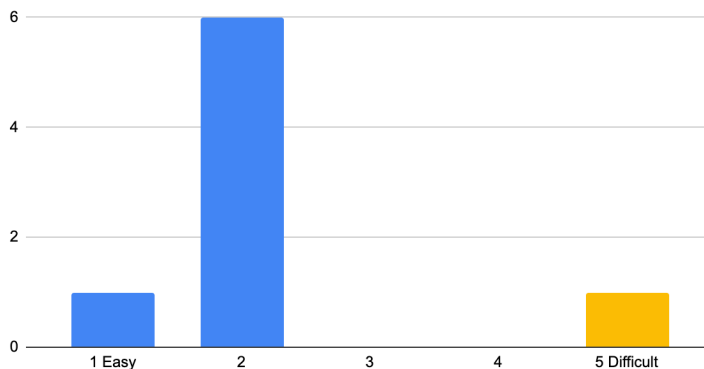
Although general ratings were positive, some notable points of friction and frustration were observed for some editors. Many of these were input-related. For example, P3 and P6 experienced difficulty inserting links, modifying text, and felt annoyed having to correct all the numbers in the MT outputs which weren't translated automatically. P2 similarly faced challenges with links and reflected about the possibility of machine-assisted links: "If I could write it in Bengali and then the link just appears for me [without manually searching], that would be great!". Meanwhile, P4 would frequently type a character once with the default Samsung keyboard, but have it appear many times; they did not understand why this was happening. Relatedly, P1 discussed how in some cases formatting problems would be handled afterwards in source editor.

"It's difficult to put the cursor in the right place to add punctuation." - [P2]

"If you could fix the font problems and the MT options, then anyone could use Section Translation easily" - [P6]

"The numbers they don't translate very well, usually they remain in English." - [P6]

"MT is very poor at figures, I have to translate them every time." - [P8]

As for input preferences, two participants (P1 and P7) were observed using GBoard, and both P5 and P6 preferred to use a phonetic, predictive input option.

In no sessions were editors observed using voice-to-text options. At the end of the session, when asked if they used such options, answers varied. Both P7 and P4 said they simply hadn't ever considered the option. P2 reported that punctuation was problematic, and P6 expressed a simple preference to type.

For some participants, what presented a problem was the general feeling of being slowed by having to work with the visual editor interface (P1, P2, P8). All three of these editors were experienced and accustomed to source editor; they all expressed the wish to be able to toggle between the two editors. In the words of P2, "If I could have source editor as an option, that would be great". P7 also noted a request to see "advanced options", which may help traditional desktop editors build comfort editing on mobile and feeling less limited. Related to advanced options, P2 expressed the wish to see an integration with wikidata, such as an automated way of identifying when wikidata items related to contents in the section are available.

During the editing process, 4 of 8 participants were observed opening additional browser tabs. P3 and P5 opened the source language article, P4 opened a dictionary and the target language article, P6 opened Bing Translator, and P7 opened a number of the linked articles.

Although all but one editor understood that the translation process proceeded sentence-by-sentence, three participants (P1, P2, P7) preferred working at the paragraph level, or at least with 2-3 sentences at a time. Thus, similarly to what was found in earlier prototype testing, editors vary in terms of whether they prefer to work sentence-by-sentence or paragraph-by-paragraph.

"For accurate translation, I need to see the exact paragraph"   - [P1]

"If I do it one sentence at a time, it slows me down"   - [P2]

The reasons for preferring to work at the level of the paragraph varied; for some it was about accuracy of translation but for others it was about speed or the general feeling of creativity throughout the process and feeling 'unhampered' or constrained by the structure of the source language section. P7 in particular

expressed feeling restricted by having to work sentence-by-sentence. Contrastively, others like P9 enjoyed checking sentences one-by-one. This highlights that editors simply differ regarding their preference on this issue.

Also related to the nature of the workflow, 7 of 9 participants understood that the first step was to edit the title of the section and then continue to the main text. P7 skipped the title during observation and was never observed returning to it before publishing. In the case of P5, at the time of the session MT was not working, so the absence of MT outputs may have caused some of the confusion. P5 selected to edit the title and directly began inserting additional text into the title field by referencing the source section in another tab and manually translating.

Topic 4 | Key takeaways | Editing translations

- **Aside from time required to edit MT outputs, main points of friction were related to input, links, and numbers**
  Numbers in particular were noted to almost never translate automatically.

- **Experienced editors want a source editor option**
  These editors feel slowed using visual editor, and are also looking for advanced options and integrations with Wikidata in one case.

- **Participants understood that the process proceeded sentence-by-sentence**
  Moreover, relatively few issues were noted around the general workflow during the editing MT outputs process.

- **The sentence-by-sentence flow doesn't work for all editors**
  Editors were divided according to whether they had a strong preference for translating at the paragraph or sentence level.

Topic 5 | Publishing

Do editors feel confident enough about their translations to complete the task by publishing?

Although most participants ultimately selected an article for the purposes of the session continuing (due to difficulty finding a relevant article from the list of suggestions), all but one participant reached a point during the session when they felt ready to publish. The only participant who did not was an editor for whom MT was temporarily unavailable at the time of the session.

Beyond feeling confident enough to publish, 7 or 9 participants actually published during the session. Upon publishing, P2 noted, "it gave me the assurance that it added my section," which the translator really liked. Again, for one participant MT was not working, and the other participant who didn't publish needed more time than available during the session.

After publishing, the most common reaction to what was needed most immediately was (1) adjusting the order of the new section to ensure proper order[3], and (2) reviewing and fixing any errors or issues that weren't easy to do on a mobile phone (for those with desktop access).

"It went at the bottom, but it should be in the middle. I will have to fix it manually with source editor [by copy/pasting]. That slows me down."   - [P2]

"I usually use source editor after I'm done with the translation; I use source editor to correct the translation"   - [P3]

P2, P3, P6, and P8 all discussed how the most urgent to-do was to ensure the section was placed in the proper order. To do so, P6 switched to desktop mode on their phone, and P2 and P8 noted how they would use source editor on their laptop to accomplish the task of recording the section via copy/paste. When discussing the next step of copy/pasting the section into the correct order, P8 noted, "for that I have to use my laptop."

---

[3] At the time of testing, the new section was universally placed at the end of the article.

Other than section reordering, P1, P6, and P8 also discussed reviewing the new section to check for mistakes and further edit the section. P1 already had some corrections in mind as they'd noted many extra spaces which were cumbersome to correct on their phone; this editor planned to do these particular revisions from a laptop. For some participants, it was clear that they viewed Section Translation as a step in a larger process, followed most immediately by a review. P1 expressed their wish for a way to categorize and know which articles are created with translation as it would be easier for others to review and correct any issues.

**Topic 5 | Key takeaways | Publishing**

- **7 of 9 participants published successfully**
  Most editors described publication as one step in a larger process, followed most immediately by reviewing and making further revisions.

- **Most participants wanted to immediately correct section ordering**
  Editors liked getting confirmation that their section had been added, but immediately wanted to correct the ordering. This was not an easy copy/paste task on a phone.

Topic 6 | General feedback

In addition to those already mentioned, two additional rating tasks were posed to participants immediately following their use of Section Translation. These asked for feedback on the overall experience and a comparison with their typical process.

First, as shown below, participants were asked to rate the translation and editing process *overall*. Just over half responded 'easy', one was more neutral, and one rated on the 'difficult' end of the spectrum, due to some of the points of friction already discussed in this report.

Overall, how would you rate the translation and editing process?
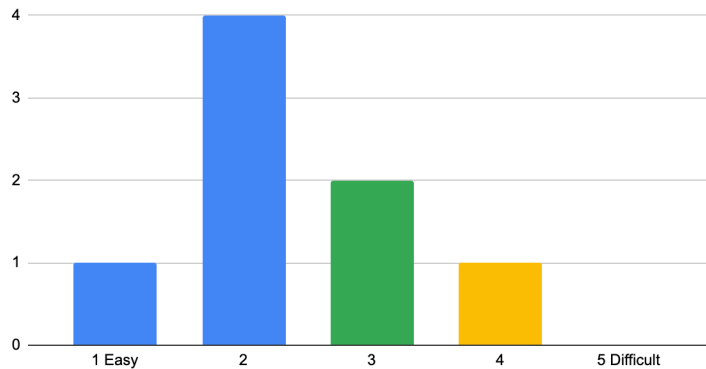
Next, we asked participants to compare their experience with Section Translation with that of their typical process of translating Wikipedia articles. This was left intentionally broad, and based on what we learned about participants typical translation processes, can be interpreted as including a variety of workflows - from frequent Content Translation use to more manual processes leveraging only visual or source editor and a variety of other tools, such as dictionaries and/or machine translation options such as Bing Translator, among others. As shown below, about half of participants responded that Section Translation was 'easier', while the other half was split between 'about the same' and 'more difficult'.

How did this process compare to your typical process of translating Wikipedia articles?
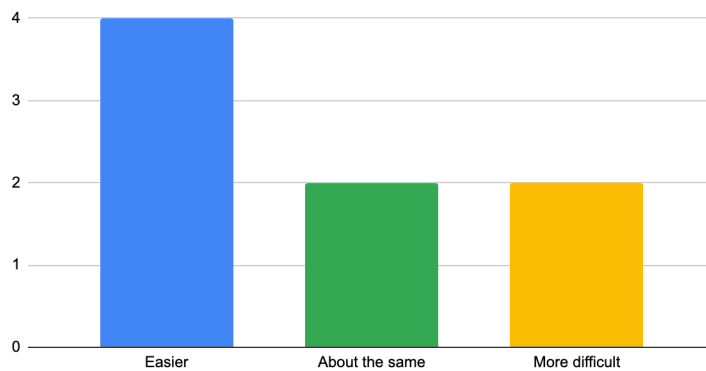
In addition to general perceptions of their experience, we also gathered information about what participants enjoyed *most* and *least* about their experience using Section Translation. Here are a few highlights from the interview portion at the end of each session.

What did you like *most* about the editing process?
- Easy navigation and interface (P1, P2, P6, P7, P8)
- Line-by-line translation (P3, P5)
- Having everything in one tab (P4, P9)

What did you like *least* about the editing process?
- No source editor (P1, P2)
- Forced to work at the sentence level (P1, P7, P8)
- Number of screens (P2)
- No search available to identify articles (P4, P5)
- Lack of Bing Translator option (P5)

It's noteworthy that the sentence-by-sentence workflow was both noted as a feature participants enjoyed, but also listed as one of the least liked aspects by other participants. This underscores how there is no easy one-size-fits-all solution to the size of chunks a tool forces translators to work with. This theme also notably surfaced in previous Section Translation research.

One pain point for participants with extensive experience using CX was that Section Translation was described as having many screens, or page loads. For example, P2 discussed this relative to internet speeds, noting that it caused a great slow down when comparing the Section Translation and CX experience. There are, "too many pages I have to change," they noted

Notable aspects of the interface that were specifically referenced as desirable included the nature of having everything available in one tab, and having the source language available at the top of the screen when editing the MT output. Both of these were important for editors due to the time they saved during translation.

"I don't have to open different tabs to translate, so in my opinion that was really easy" - [P4]

"The missing portions [sections] can be inserted quite easily with Section Translation" - [P8]

"I open in one tab the English article and the Bengali article in another; then I also open Google Translator or Bing Translator in another tab if I need it [describing usual, non-Section Translation process]" - [P4]

Topic 7 | Device patterns and the unfinished nature of translations

In the discussion of *Topic 5 Publishing*, we noted how many editors described next steps that involved returning to the published section and article - not only to modify the order of the newly added section, but also to review for typos and errors. Through conversations with participants, we can note that publishing the section with Section Translation (or article with CX for that matter) is not perceived as marking the work as complete. We heard many describe how they may return to edit the article or section further using source editor on another device. Editors may even purposefully select to leave the translated passages as incomplete in some way because Section Translation/CX are one aspect to their editing workflow. For example, P6 noted how they use translation to create new content, sometimes only creating section names and the basic structure first. They return later to complete the article more fully. In a similar spirit, they noted, "If I think the link will be added later, I may just leave it [even if it's broken]".

The theme of viewing translation as part of a larger workflow also relates to device switching for editors with access to multiple devices. As noted above, many participants said they'd correctly order newly added sections using source editor on their laptop. For those with both smartphones and laptops, editors consistently describe using phones for making 'smaller' edits, fixing typos, editing references, and performing other tasks when a laptop may not be available, such as if they're traveling. Editors who tend to work more from laptops also expressed a desire to use Section Translation from their laptops. For example, P8 said, "I would love to

start translating a new article on mobile. At the same time, I would like to use this feature [Section Translation] on a laptop."

Topic 8 | Bing Translator

Around half of participants expressed a desire to use Bing Translator due to it producing higher quality outputs for the English - Bengali pair. Participants consistently described Bing as superior to Google when it comes to sentence structure (e.g., P4, P6). P6 described using Google Translate (GT) only for individual words and said they were encountering meaningless passages in the GT sentence-level outputs. This editor was observed copy/pasting text out of Section Translation into Bing Translator and back because that manual process was perceived as easier than just defaulting to Google's MT outputs. With Bing Translator, they felt they only needed to modify a few words.

Key takeaways, recommendations, and next steps

Top 10 takeaways

1. **Translators want to search for specific articles**
   Without the ability to translate articles of specific interests they perceive as new translation opportunities, editors face a near hard stop with Section Translation.

2. **Editors' workflows begin at the source wikis**
   Regarding Section Translation entry points, translators generally begin their work at a source wiki where they are browsing and searching for articles that match topical interests.

3. **Section selection was relatively seamless**
   Overall, section selection presented fewer points of friction when compared to article section.

4. **Section selections must provide a 'core' content part of the article**
   In order to present a compelling translation opportunity, missing sections must pertain to core content of the article and topic. Editors were less likely to feel compelled to translate closing sections perceived as peripheral, such as 'references' and 'external links'.

5. **Interactions with the quick tutorial were very brief**
   Few if any questions arise for editors at this point in the process, so the frequency with which these are shown can be reduced to avoid annoyance.

6. **Aside from time required to edit MT outputs, main points of friction were related to input, links, and numbers. Bengali editors want Bing Translator support.**
   Numbers in particular were noted to almost never translate, and Bing is frequently preferred over Google for providing semantically- and syntactically-superior outputs.

7. **Experienced editors want a source editor option**
   These editors feel slowed using visual editor, and are also looking for advanced options and integrations with Wikidata in one case.

8. **The sentence-by-sentence flow doesn't work for all editors**
   Editors were divided according to whether they had a strong preference for translating at the paragraph or sentence level. Some listed it as one of their most liked parts about the editing process, whereas for others it made their list of top dislikes.

9. **The majority of participants published successfully**
   Participants said they could work faster with everything in a single tab with Section Translation. Still, editors described publication as one step in a larger process, followed most immediately by reviewing and making further revisions.

10. **Most participants wanted to immediately correct section ordering**
    Editors liked getting confirmation that their section had been added, but immediately wanted to correct the ordering. This was not an easy copy/paste task on a phone.

Recommendations

There are a number of recommendations stemming from the key findings presented on the previous page. A Phabricator ticket is being used to track various lines of work stemming from research findings: https://phabricator.wikimedia.org/T279064. This ticket can be used to view and track ongoing work, including tickets related to some of the items in the list below.

10 key recommendations include:

1. Support search to identify a section/article to translate
2. Improve section mapping
3. Improve suggestions
4. Support Bing Machine Translation
5. Adjust when to show the quick tutorial
6. Ensure that core content sections are available when suggesting articles
7. Improve published section ordering by implementing a first best guess and providing a mobile-friendly way of adjusting the order of sections
8. Improve discoverability of Section Translation, including entry points at common source wikis (ongoing design research work: T237385)
9. Discuss the complexity and relative priority of providing source editor for Content and Section Translation
10. Continue exploring ways of supporting both translators who prefer working at the sentence and paragraph level

Next steps

As the number of wikis where Section Translation is available increases, subsequent testing should be used to understand the specific needs of different editing communities that span linguistic, cultural, and geographic divides. Such research should be timed to correspond with the completion of additional development efforts stemming from the recommendations from earlier reports, such as those noted above. Follow-up testing should also specifically attempt to validate these design and technical improvements.