

# content translation survey feedback

Understanding Content Translation (CX) newcomer experiences through survey feedback collected at early 2020 edit-a-thon events in India.

Eli Asikin-Garmager  
Language Team, Wikimedia Foundation  
*[easikingarmager@wikimedia.org](mailto:easikingarmager@wikimedia.org)*

---

[Background](#)

[Results](#)

[Next Steps](#)

**T**ranslation is essential to a world in which everyone can share freely in all the world's knowledge.

Although Wikipedia is accessible in approximately [300](#) languages, content across different language Wikipedias differs significantly, especially in terms of the number of articles available. In efforts to close these content/knowledge gaps, [Content Translation \(CX\)](#) has now been used to create [more than 600k](#) Wikipedia articles. By automating boring tasks, requiring less content expertise, and not requiring source code editing, CX is helping to grow the number of contributors in small wikis.

Currently, most user feedback on the tool is received by the Language Team through [talk page messages](#) and direct contact from active, vocal editors. This feedback has been very valuable, and at the same time there is opportunity to create new routes for feedback on the tool; routes that increase the diversity of voices represented, improve consistency in feedback methods and metrics, and make it easier for community members to help collect feedback on the tool. In particular, collecting feedback

from new CX users is important for understanding how to grow the Content Translation user base.

The goal of [this project](#) was to develop and iterate upon an easily-customizable survey that can be used to more easily, quickly, and reliably collect feedback from a more diverse pool of CX users and ensure new user voices are adequately represented. This report presents results from the first phase of the project.

## Background

### How do we ensure *new* CX user voices are represented and promote an open feedback loop?

During Phase 0 of the [CX Feedback Survey project](#), the goal was to develop an initial survey based on the Language Team's current [backlog](#) of research questions. Survey questions and response options were based on previous interviews with users from the Bengali and Javanese Wikipedia, as part of the [Section Translation](#) design research project.

The survey was piloted in coordination with CX edit-a-thon events in India, and addressed the following categories:

- General usability - what are user preferences, pain points, and barriers to use?
- Perceptions of machine translation incorporated in CX
- CX value propositions and perceptions
- New CX user feedback and initial interactions

Prior to distribution, the survey was translated from English to Punjabi and Hindi, a decision made in coordination with the edit-a-thon event organizer in light of language abilities and preferences among participants. After review by the Language Team, the survey was built out in these three languages using Qualtrics.

In addition to the different language versions of the survey, there are [3 versions of the survey](#). These are intended to collect feedback from participants at different touch points during the edit-a-thon events. **Version 1 was administered after participants completed Day 1 of the event.** Version 2 is intended to be administered after participants have published 5 articles via CX or after 15 days, whichever comes first. Finally, Version 3 is administered after the participants' language community has reached their goal (target number of articles) for the event.

On 2 March 2020, version 1 of the survey was shared with participants by an edit-a-thon organizer in India, whose group for

the event used CX to translate English articles on Cyber Crimes ([Syberthon 2020](#)) to a number of Indic languages, including Punjabi, Hindi, Bengali, Gujarati, Tamil, Telugu, Malayalam, and Marathi.

**This report shares results from Version 1 of the survey,** which was administered and completed just before the [2019-2020 Coronavirus pandemic](#) caused the edit-a-thon events to pause. As of 29 April, results from version 2 are anticipated in roughly 2 months time. Events are still proceeding online with a slower pace.

## Results and discussion

As noted, this version of the survey was administered after the first day of the edit-a-thon event, and offered in English, Hindi, and Punjabi. We received 77 responses to the English version, 1 response to the Hindi version, and 0 responses to the Punjabi version. Before analyzing responses and results, responses from any participants who had completed less than 25% of the survey were removed. After removing data from these 10 participants, this left responses from a total of 67 respondents, 55 of which completed 100% of the survey. Without a clear rationale for excluding data from 12 respondents who may not have

completed the survey in full (e.g. 68%, 82%), these data were included in the analysis.

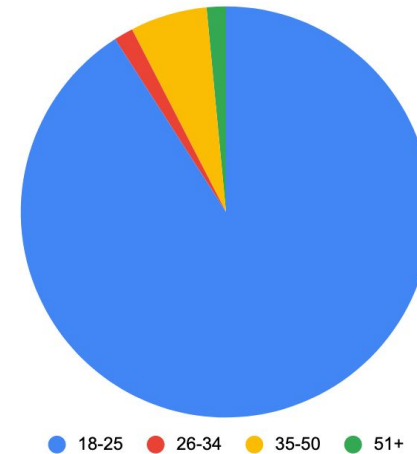
## Respondents

In this section, we examine respondents based on self-reported data provided in the survey around basic demographics and prior experience with Wikipedia and CX. **Overall, the users sampled for this survey were young, new users. Most had heard of, but not previously used, Content Translation (CX).**

The sample skewed towards *young users*. As shown in *Figure 1*, the vast majority (91%) of participants were between the ages of 18 and 25. Those whose ages were between 35 and 50 represented an additional 6%, and the remaining 4% of users were divided equally between the 26-34 and 51+ age groups.

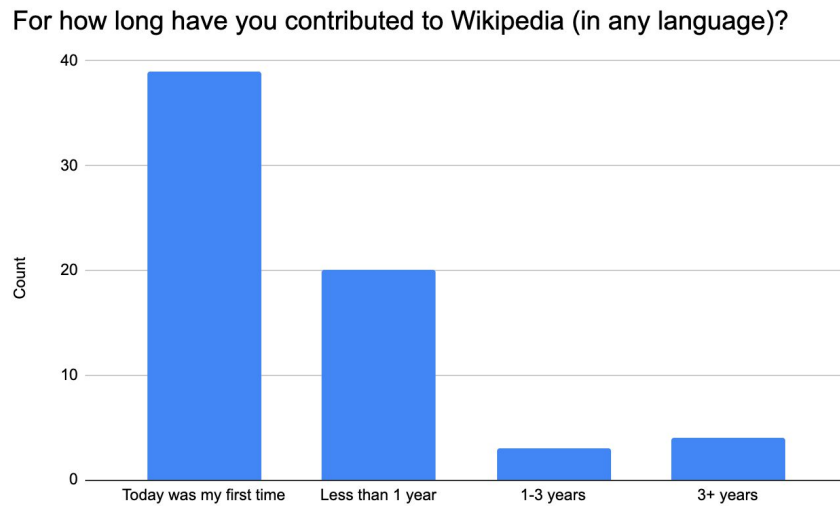
Fig 1. Respondent age

Which best describes your age?



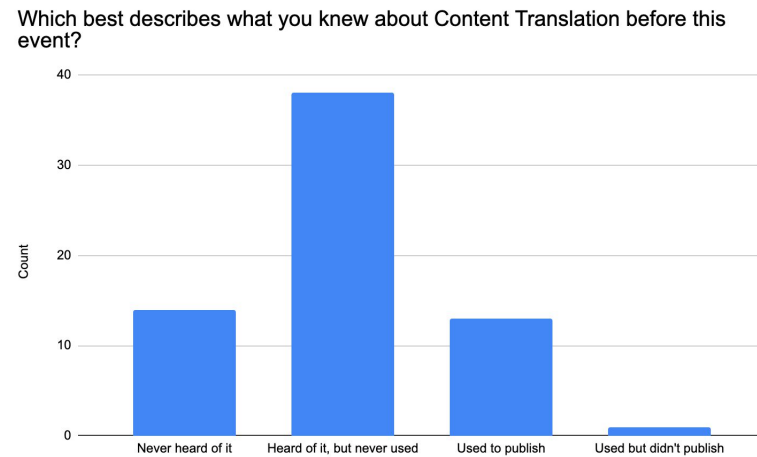
The sample for this survey also provided strong representation of *new Wikipedia users* - new to both Wikipedia and Content Translation. As shown in *Figure 2*, the majority (n=39, 59% overall) of respondents had not previously contributed to Wikipedia, and most who had came with editing histories of less than one year (n=20, 30% overall). A small number (n=7, 11% overall) had contributed to Wikipedia for more than one year at the time of the survey.

Fig 2. Wikipedia contribution history



Given that most respondents were new Wikipedia contributors, it follows that the majority were also new Content Translation users. As shown in *Figure 3*, prior to the edit-a-thon event (during which the survey invitation came) 58% had heard of Content Translation (hence their potential interest in the event), but never used it. Another 21% had never heard of Content Translation. Only 20% had actually used it to publish a translation.

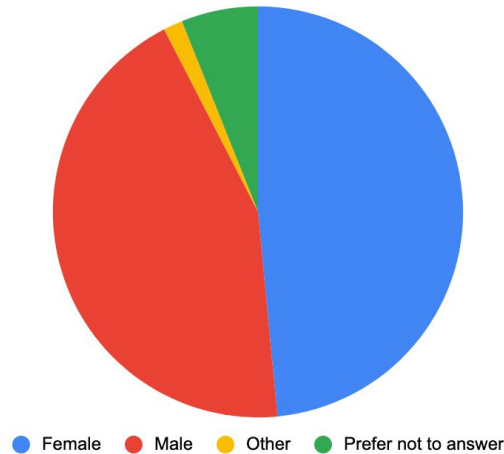
Fig 3. Previous experience with Content Translation



Finally, as for gender representation, there was a roughly 50/50 distribution of male/female users. As shown in *Figure 4*, 48% of respondents self-identified as female, and 44% as male. Another 2% responded other, and 6% opted to skip this question.

Fig 4. Respondent gender

To which gender do you most identify?

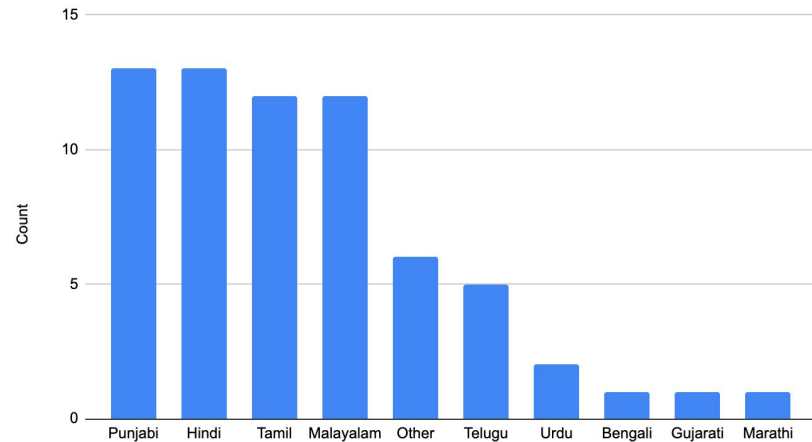


## Language Community Representation

While a number of different language Wikipedias were represented among respondents, 76% of respondents were roughly equally distributed among four different Wikipedia language communities: *Punjabi* (20%), *Hindi* (20%), *Tamil* (18%), and *Malayalam* (18%).

Fig 5. Language Community Representation

To which Wikipedia did you contribute today?



Telugu was the next largest group, corresponding to 8% of respondents. Urdu, Bengali, Gujarati, and Marathi each had between one and three respondents, and another six respondents were distributed across Kashmiri, Nepali, Manipuri, Assamese, and Mizo.<sup>1</sup> A summary of this breakdown is provided in *Figure 5*.

<sup>1</sup> Mizo is an incubator wiki, and thus CX is not currently available. It is not clear if the 1 respondent who reported that they contributed to the Mizo Wikipedia did so without use of CX, or if they contributed to another wiki using CX.

## Content Translation User Preferences - Likes

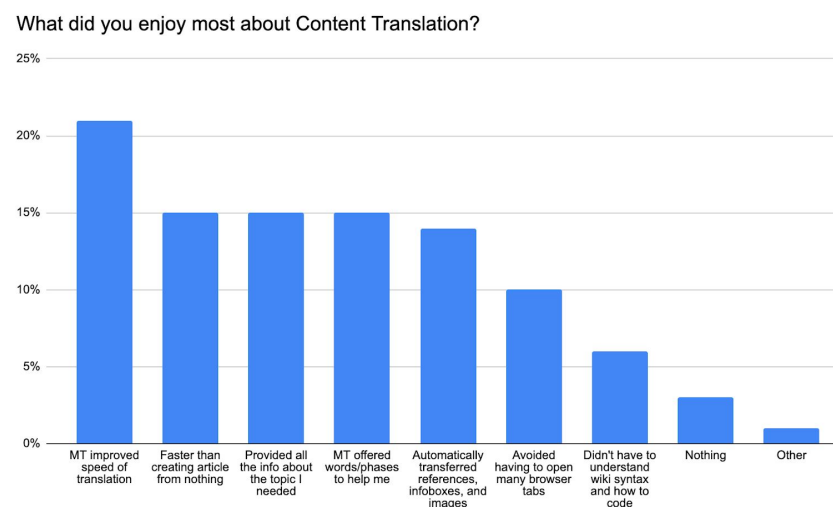
As noted earlier, the sample for this survey was largely composed of *young, new* Content Translation users, most of who had *not* previously contributed to Wikipedia in any language. This section presents overall preferences around CX. Respondents were able to select up to three responses for this question.

When asked what they enjoyed most about CX, the most frequently selected response among respondents was that the *'machine translation (MT) improved the speed of translation'*. The top five most frequent responses were:

1. **The MT improved the speed of translation**
2. CX was faster than creating the article from nothing
3. CX provided all the information about the topic I needed
4. The MT offered words/phrases to help me
5. CX automatically transferred references, infoboxes, & images

As shown in *Figure 6*, the other four most frequently selected responses (#2-4 in the list above) were roughly equal in their distribution. Percentages in *Figure 6* show the relative frequency of a particular response type; they don't reflect the percentage of respondents giving a particular response.

Fig 6. What respondents enjoyed most about CX



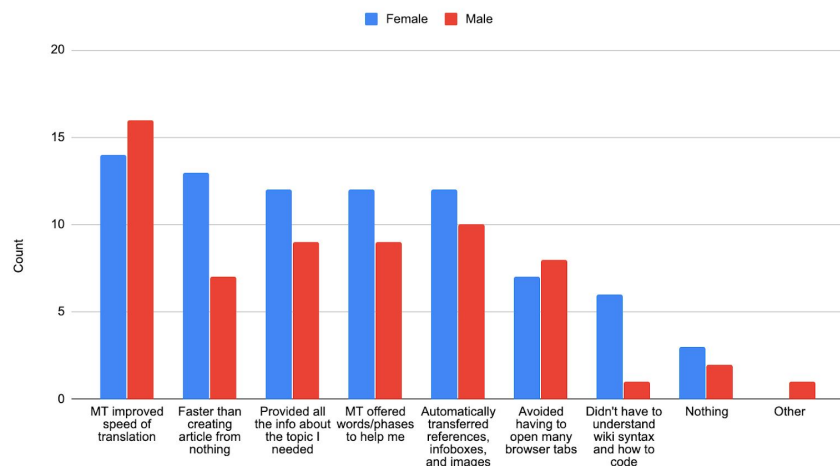
Worth noting is the fact that two of these top responses reference *speed*, suggesting that users value the time-saving nature of the feature.

Given the number, and roughly equal distribution, of respondents by gender, we can briefly examine any preference differences across these groups. Breaking down responses to this question by gender (*Figure 7*), it is clear that some response options patterned quite similarly regardless of gender. Indeed,

while *Figure 7* shows some potential tendencies for variation, none of these differences were statistically significant.<sup>2</sup>

*Fig 7.* What male/female respondents enjoyed most about CX

What did you enjoy most about Content Translation?



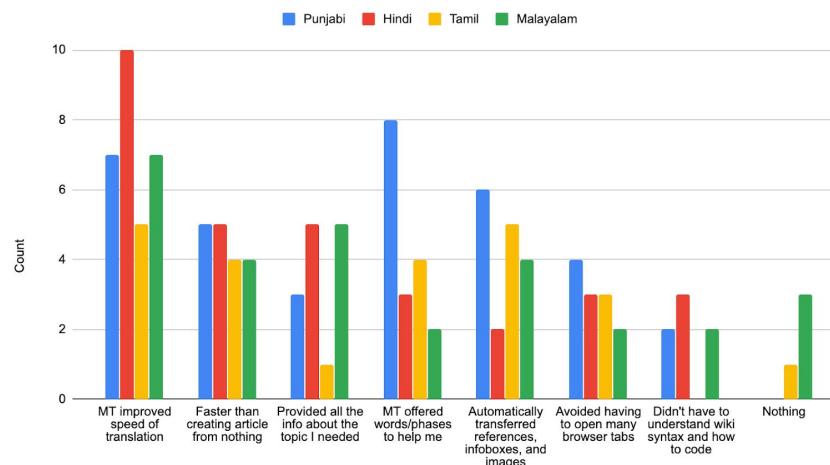
While sample size is limited in terms of examining these preferences by language community, given varying characteristics across languages (e.g., MT quality and wiki culture,

<sup>2</sup> A chi-square test was used except where the expected value for any group was less than 5, in which case a Fisher's exact test was used. P values ranged from .25 to .81. The only two comparison that approached marginal significance was 'CX was faster than creating an article from nothing' ( $X^2(1, N = 20) = 0.14, p = .25$ ) and 'I didn't have to understand wiki syntax and how to code' ( $p = .11$ ; 2-tailed Fisher Exact test).

among others), *Figure 8* presents responses to this question broken down across the four wikis for which we had the most respondents (Punjabi, Hindi, Tamil, and Malayalam). No statistical tests were performed on these data due to sample size concerns.

*Fig 8.* Preferences by language community

What did you enjoy most about Content Translation?



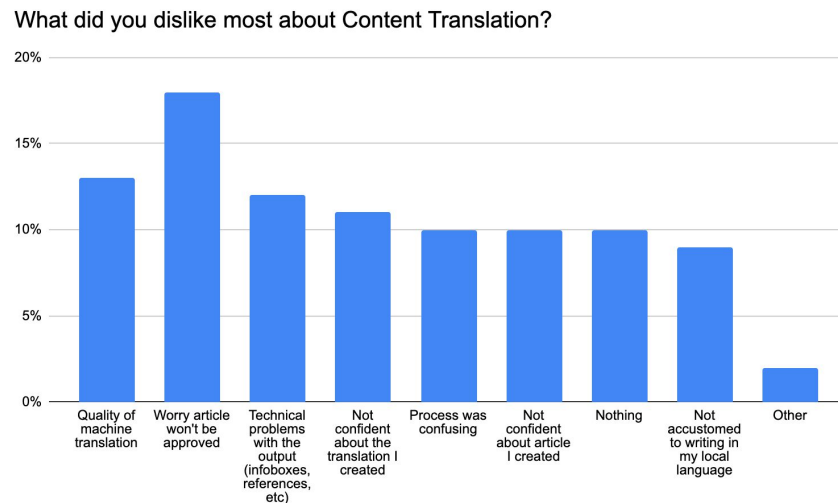
While offered here as tentative results, it's easy to note that there are potentially some wiki-by-wiki differences in *Figure 8*. A greater sample size from a limited number of different language wikis would allow us to test for statistical significance and ensure a representative sample.



## Content Translation User Preferences - *Dislikes*

Respondents were also asked what they disliked most about Content Translation. Respondents were prompted to select up to three options. Overall, preferences were quite spread across the responses offered, more so than in the case of what respondents enjoyed the most.

Fig 9. What respondents disliked most about CX<sup>3</sup>



<sup>3</sup> Again, these percentages show the overall frequency of any given response being selected; not the percentage of participants selecting that response.

The most frequent response to this question was that respondents, 'worried the article wouldn't be approved.' The other responses near the top in terms of frequency included 'quality of the MT' and 'technical problems with the output (infoboxes, references, etc).'

Just marginally outside of the top three responses was lack of confidence about the translation the respondents had created and general confusion about the process and lack of confidence around the article as a whole. So, while technical considerations such as MT quality and technical problems were important dislikes, general lack of confidence around the translation, article, and approval process was top of mind for respondents.

Respondents were presented with a follow-up open-ended question if they selected 'the process was confusing' or 'they lacked confidence in their article'. When asked what about the process was confusing, respondents gave responses suggesting that confusion stemmed from different types of content they worked with, as well as the publishing process. For example, event organizers reported that it was very common for contributors to skip over translation of the article title. They'd often begin directly with the article content, and mentors had to provide frequent reminders around not forgetting to translate the article title. Other word-for-word responses included:

- 'Info box'
- 'Confusion in context'
- 'About the publishing'

The second follow-up question was asked to any respondents that reported lacking confidence in their article. Open-ended responses varied, but are suggestive that confidence was eroded to an extent by encountering warning messages around MT and writing in their local language. Word-for-word responses included:

- 'I was written in my own language (Kashmiri) but Wikipedia is not supporting Kashmiri language'
- 'Because even after writing a whole paragraph in my own language, Wikipedia said its 100% machine translation'
- 'The words i used and writing in vernacular language was also a difficult one'

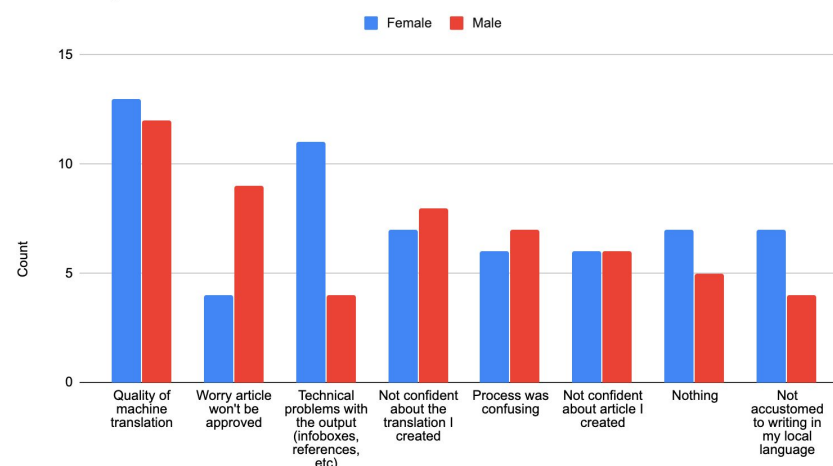
Related to the first respondent quote is the general pattern of how contributors in these wikis are often faced with topic-specific technical vocabulary, for which translation into the local language is difficult. In many cases respondents use transliteration to deal with these lexical challenges.

While no statistically significant gender difference arose around what respondents enjoyed most about CX, we briefly examine the role of gender as it relates to CX dislikes. As shown in *Figure 10*, there was a tendency for males to more frequently

report worrying more about their article being approved than for females, but this difference was not statistically significant ( $X^2(1, N = 13) = 2.35, p = .13$ ) In fact, there were no statistically significant differences (other P values ranged from .10 to 1). Only one male/female difference was marginally significant: females more frequently reported technical problems with the output of CX (e.g., infoboxes, references, etc.) ( $X^2(1, N = 15) = 2.74, p = .10$ ).

*Fig. 10.* What male/female respondents disliked most about CX

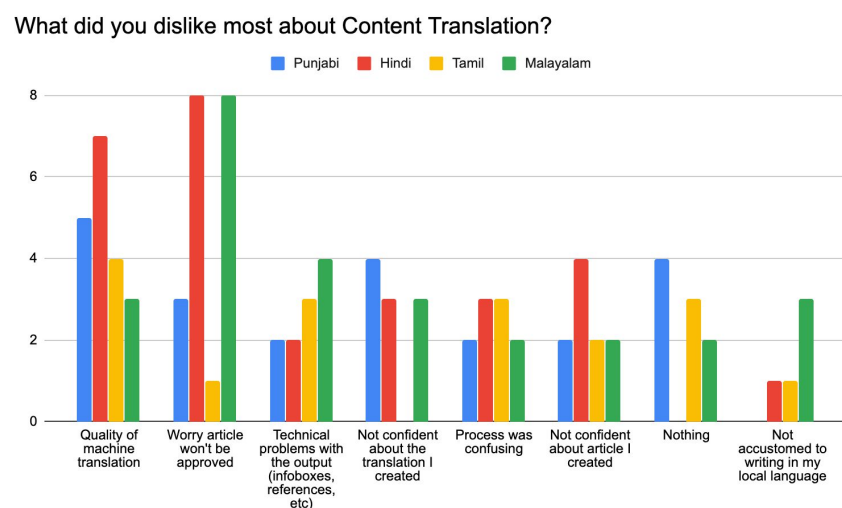
What did you dislike most about Content Translation?



As with the previous question, while sample size was limited in terms of examining dislikes by language community, given varying characteristics across languages (e.g., MT quality,

wiki culture, etc), *Figure 11* presents responses to this question broken down across the four wikis for which we had the most respondents (Punjabi, Hindi, Tamil, and Malayalam). No statistical tests were performed on these data due to sample size concerns.

*Fig 11.* Dislikes by language community



While offered here as tentative results, it's easy to note that there are potentially some wiki-by-wiki differences in *Figure 11*. For example, Tamil Wikipedia editors showed less concern about their article being approved compared to Malayalam and Hindi contributors. Somewhat surprising, given the size of Hindi

language training corpora compared to Malayalam corpora, more Hindi contributors expressed concern/dislike around the quality of machine translation. It might be that as the quality of machine translation increases, so do expectations.

Again, *Figure 11* is offered as a preliminary glance at some potential variation by language community, but a greater sample size is needed to make more conclusive statements.

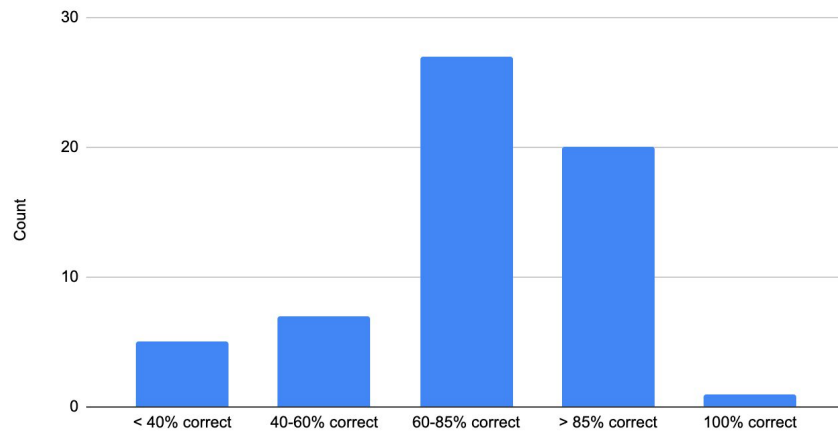
## Perceptions of machine translation

Next, respondents were asked to rate the machine translations produced by CX during their experience with the tool.

As shown in *Figure 12*, the majority of respondents rated the MT outputs as 60-85% correct, with the second largest group of respondents rating the output as more than 85% correct.

Fig 12. Perceptions of machine translation

Overall, how would you describe the machine translations produced automatically by the tool?



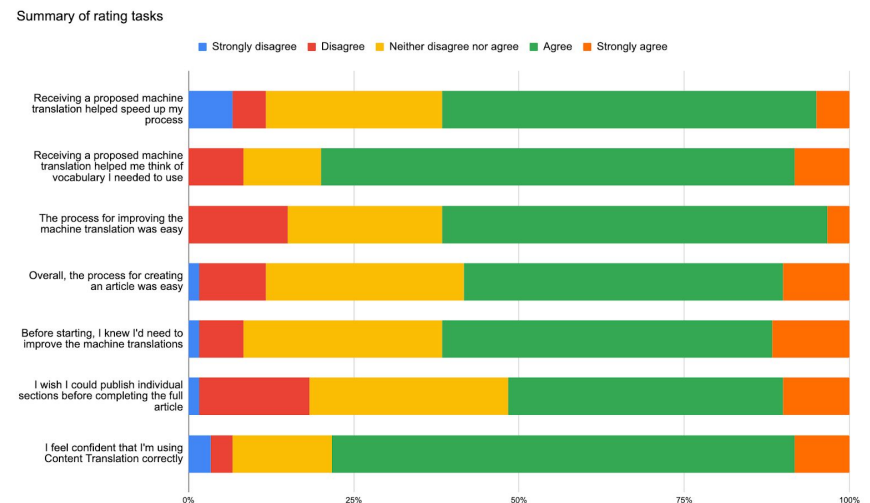
## Level of agreement with statements about CX

The last part of the survey asked respondents to state their level of agreement with a series of seven statements about Content Translation. An overview of results is provided in *Figure 13*, and the seven statements included:

1. Receiving a proposed machine translation helped speed up my process
2. Receiving a proposed machine translation helped me think of vocabulary I needed to use

3. The process for improving the machine translation was easy
4. Overall, the process for creating an article was easy
5. Before starting, I knew I'd need to improve the machine translations
6. I wish I could publish individual sections before completing the full article
7. I feel confident that I'm using Content Translation correctly

Fig 13. Level of agreement with statements about CX



For **statements #1 and #2**, we note respondents were less likely to both disagree and strongly disagree with the first compared to

the second, suggesting CX provides a stronger benefit to new users in how it offers needed vocabulary. A total of 62% of respondents agreed with the first statement, whereas 80% agreed with the second.

While no respondents strongly disagreed, some provided neutral and negative responses to **statement #3**, 'the process for improving the translation was easy'. If a respondent did not respond favorably to this statement (i.e., agree or strongly agree), they were presented with a follow-up open-ended question asking what was difficult about improving the machine translation. Most responses centered around MT output for vocabulary options. Based on the responses provided, most difficulty was lexical, not syntactic, in nature. For example, one respondent wrote that they, "couldn't get proper words when translating", and another commented about, "words which get translated are not as per the requirements".

Similarly related to ease of use was **statement #4**, which asked respondents to agree/disagree with how easy the overall process was. For respondents not agreeing or strongly agreeing, we received 5 responses to a follow-up question about what the most difficult part of the process was. The majority of responses noted challenges around the translation of certain vocabulary (particularly technical terms), one respondent specifically noted syntactic problems with the MT, and yet another mentioned the

publishing process. Word-for-word responses included the following:

- 'Search words in the language to be translated, like some technical words'
- 'Publishing'
- 'Changing the sentence, and changing grammar mistakes'
- 'Spelling'
- 'Translation was not accurate'

Next, **statement #5** responses were interesting because it is clear that not all respondents realized they would need to improve the machine translations produced in the CX tool. More specifically, a total of 39% of respondents did not agree to the statement (i.e., disagreed or were neutral). In other words, new CX users don't all have the same assumptions around the role of MT (and their role in relation to MT). They don't all realize what their role will be relative to the MT output, and don't assume they'll need to improve the MT offered.

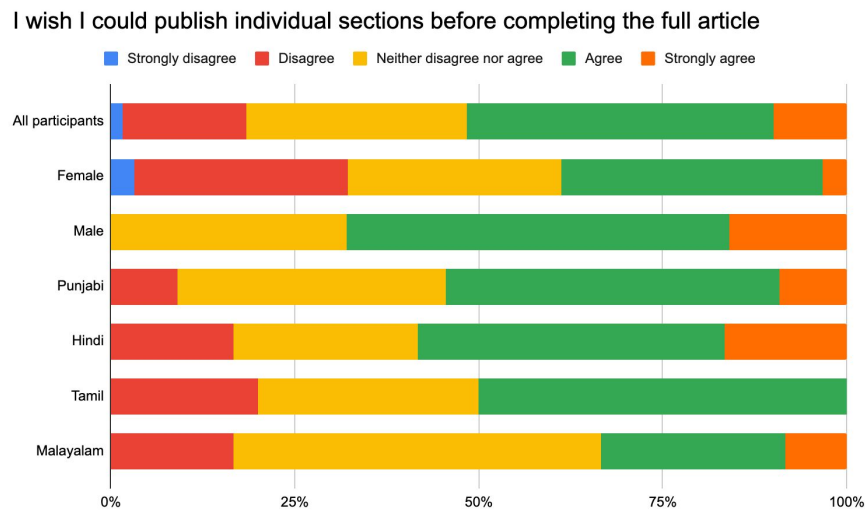
Before moving to statement #6, let's briefly note that responses to **statement #7** (confidence about using CX correctly) were overall positive, with the majority (78%) of respondents agreeing or strongly agreeing that they felt confident they were using CX correctly.

Finally, **statement #6** asked respondents about the importance of being able to publish individual sections of an

article. Just over half (52%) agreed to some degree that they wished they could publish individual sections. So, while the majority agreed, around 30% were neutral and 19% disagreed. Given upcoming Language Team decisions around how Section Translation will be implemented, it's worth taking a closer look at responses to statement #6

Figure 14 breaks down responses to statement #6 both by gender and language community (for the four most-represented language communities in our sample).

Fig. 14. Preferences around publishing individual sections



Examining gender differences first, it's striking that no male respondents disagreed to any extent with this statement, suggesting males are more favorable than females to Section Translation. As for differences by language community, it is worth reiterating that sample sizes are quite small when breaking down the data by community. Further data should be collected to help understand language community variations in receptiveness to Section Translation, but it's worth noting that Malayalam respondents were slightly less likely to agree with the statement about wishing they could publish individual sections. Similarly worth noting is that no respondents from any of these four language communities responded 'strongly disagree' to the statement about section translation. Much of the lack of agreement fell in the 'neither disagree or agree' category, suggesting respondents didn't feel strongly about it and/or were not sure exactly how and why one might do this.

## Discussion and key takeaways

While this preliminary report covers only version 1 of the Content Translation survey, there are a number of takeaways, both for our understanding of how new users interact with CX as well as for how we might proceed in the future with regard to data collection. This section presents 10 highlights of what we've learned so far from these data.

## Survey administered in coordination with edit-a-thon events

1. A survey administered in coordination with edit-a-thon events was an effective way of collecting feedback from new CX users. It also provided for a balanced sample in terms of gender and representation from many language communities.
2. Administering in coordination with an edit-a-thon allowed us to collect feedback on general usability and perceptions, but not around discovery of the tool (since it was a guided introduction). Administering via the live interface would provide a better opportunity for collecting feedback around entry points.

## New CX user preferences, pain points, and opportunities

3. Speed matters. When examining what users enjoyed most about the tool, features that benefit speed of translation and article creation came in at the top.
4. Vocabulary support options are highly valued as much as the overall option of having MT available is.
5. Top dislikes about the tool were related to concerns/worry about users articles being accepted (not deleted), and technical issues with different content types (infoboxes, etc...)
6. To build new user confidence, we can ease the process of publication and help build confidence that users articles will survive

7. Nearly 40% of respondents did not realize that they would need to improve the machine translations produced in CX. There are opportunities to better surface the nature of human-MT interactions.
8. Section Translation receptiveness was around 82% overall. 52% responded directly favorable and 30% were neutral, suggesting they may not have considered and thought about the option previously to being asked about the possibility.

## Gender and language community variation

9. Minimal gender differences were noted - both around what users enjoyed and disliked the most. A few gender difference tendencies were noted, but were only marginally statistically significant.
10. Potential differences by function of the language community were explored, but a greater sample size is needed to test the statistical significance of such differences and ensure a representative sample.

## Next steps

The survey will be revised and updated, and provided as a customizable resource on [Best Practices for Content Translation Events](#).