

# machine translation meets human perception

*hindi // indonesian // arabic*

Drawing on contributions by readers of *Indonesian*, *Arabic*, and *Hindi*, the Wikimedia Foundation conducted an in-depth study of how audiences perceive machine-translated Wikipedia articles.

**T**he *Wikimedia Foundation* investigated the role that Machine Translation (MT) plays in readers' perceptions of Wikipedia articles among three diverse and under-studied language communities. This study sheds light on the potential benefits and pitfalls implied by the widespread incorporation of automatic translation into the Wikipedia reading experience.

In response to the widespread and growing use of translation technologies to expand access to Wikipedia content among language communities in which target-language articles may be limited, the *Wikimedia Foundation* worked with non-Wikipedian readers of Arabic, Indonesian, and Hindi to measure the extent to which MT affects perceptions of articles' (a) *Quality*, and (b) *Utility*, or usefulness. Participants in this study read and responded to either an MT version of an English-language Wikipedia article, or a human-written (HW) article on the same topic, in their own language.

This study contributes to the broad finding that readers of Wikipedia perceive MT and HW articles differently, and that they do so in similar ways across diverse languages. HW articles in this study were universally rated as higher *Quality*, for example, while MT articles were generally found to be similarly useful to HW articles. Wikipedia readers drawn from diverse

language communities thus seem to make meaningful distinctions between articles' informational content and articles' overall language quality, the latter of which is most affected by Machine Translation.

## **Machine Translation lowers *Quality* more than *Utility* (“usefulness”)**

- MT tends to negatively affect perceptions of both *Quality* and *Utility*, but it lowers *Quality* more.
- Although MT articles are usually rated as slightly less useful than HW, MT is sometimes found to be *more useful* than HW in cases when the translated English-language source provides information that is not available in the HW target-language article.

## **“High-quality” articles are highly informative**

- Across language groups and among both HW and MT articles, “high-quality” is associated with abundant and accurate information, rather than factors such as style or language quality.

## **Readers' beliefs (weakly) influence perception**

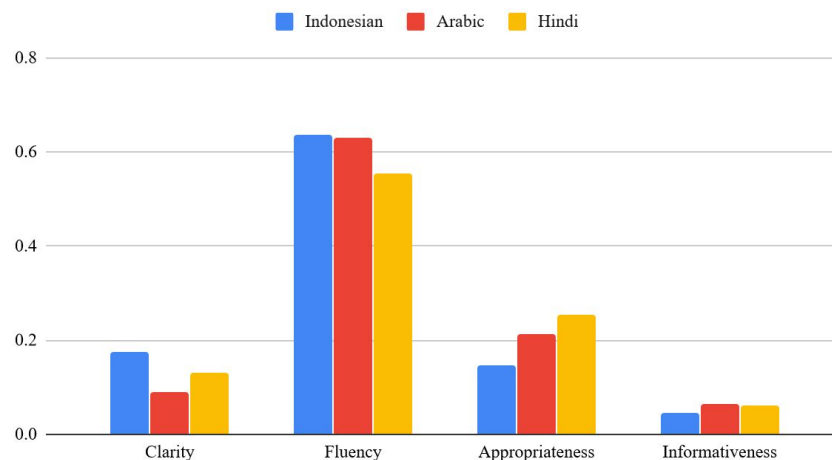
- Reported trust in Wikipedia is predictive of how useful participants find either MT or HW articles.
- Participants who use Wikipedia more frequently also have a higher degree of trust in the platform.

- Participants who say they value high-quality translation more find MT articles to be higher *Quality*.

## MT articles are limited most by their language quality

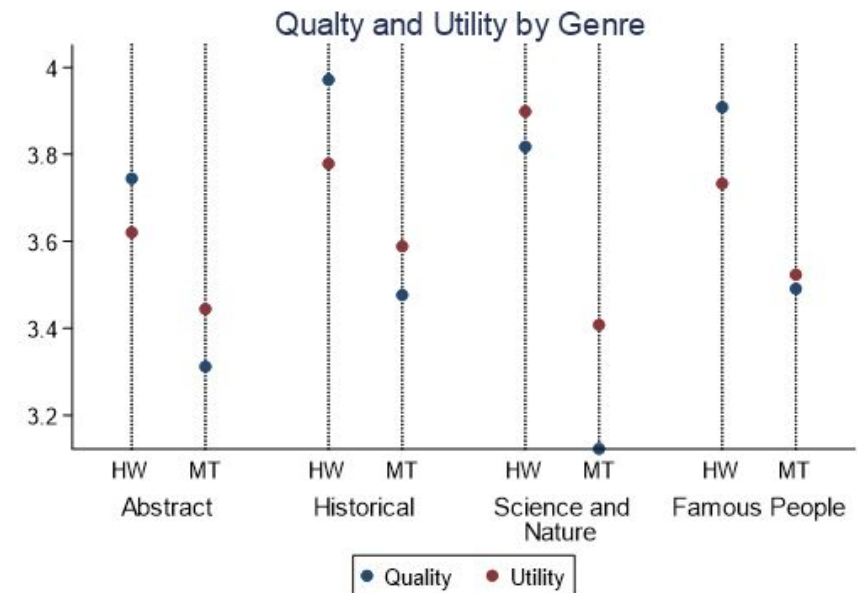
- The biggest flaws of MT articles stem from perceptions of their deficient *Fluency*; i.e., the quality of language used to compose them, rather than informational content or style.
- Perceptions of diminished *Fluency* are relatively constant across language groups:

Salient Features for Improving MT (%)



## Article genre affects perception of HW articles

- HW *Science and Nature* articles are seen as more useful than other HW genres, while HW *Historical* articles are seen as higher quality.<sup>1</sup>



- The *Quality* of MT articles is lowered more than their perceived *Utility*, regardless of genre--i.e., genre does not significantly affect perception of MT articles.

<sup>1</sup> Articles were grouped into four “genres” for comparison. The *Abstract* group included articles on topics such as “Anxiety,” “Emotion,” and “Fear,” for example.

## Different language communities may have different stances toward Wikipedia

- Arabic participants express a more nebulous stance toward Wikipedia than Indonesian participants, reporting less use of and trust in the platform.
- Arabic participants are slightly more concerned with perceived article *Fluency* than other language groups.

## Bottom lines for stakeholders

- **MT products interact with new socio-cultural landscapes in unpredictable ways.** MT output is perceived as highly formal, for example, and while this formality is desired and culturally valued by *Arabic* readers, *Hindi* and *Indonesian* readers consider excessive formality to be a barrier to access.
- MT lowers perceptions of *Quality*, but MT articles may be equally or more useful than HW, depending on informational content.

## How do readers form impressions of MT or HW articles?

Through numerical ratings and responses to open-ended questions such as “How could the translation of this article be improved?”, participants in this study helped to clarify an emerging model of Wikipedia article perception in which overall impressions of *Quality* and *Utility* are governed by distinct judgements about the text:

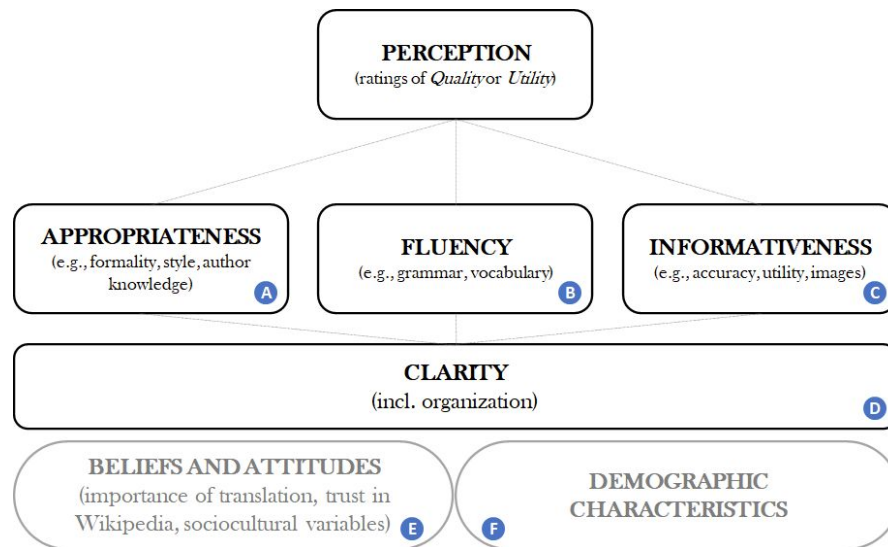
- A. **Appropriateness**, i.e., acceptability of style and formality level;
- B. **Fluency**, i.e., “correctness” of grammar, vocabulary, and overall quality of translation;
- C. **Informativeness**, i.e., abundance and accuracy of information.

Judgements A, B, and C are undergirded by the equally important judgement that readers make about an article’s

- D. **Clarity**, i.e., intelligibility and organization.

Readers’ impressions of *Quality* and *Utility* are also contributed to--albeit to a much lesser extent--by factors that readers bring with them into the reading experience, including their:

- E. **Beliefs and attitudes**, e.g., trust in Wikipedia and general value of “high-quality” translation; and
- F. **Demographic background**, e.g., preferred language of online reading and age group.



*Across language groups, Clarity and Fluency make the largest contributions to Quality, while Utility, is predicted most strongly by Clarity and Informativeness.*