Tracking the framing of politicians and news events across the multilingual Wikipedia

Mark SHUTTLEWORTH
PAN Jun
TU Zhilu
(Hong Kong Baptist University)

This paper aims to report on the results from research conducted to date on translation in this well-known online encyclopaedia and to present plans for the next stage.¹

The study of Wikipedia translation is a recent addition to the increasing body of research that focuses on collaborative and crowdsourced translation, and has hitherto focused on translation quality, professional translators’ views of Wikipedia, translation and cultural remembrance, multilingual knowledge production with reference to city-related articles and to news stories, and knowledge construction and point of view (POV). On the other hand, more work has been done on the multilingual Wikipedia within other disciplines.

Wikipedia pursues an official Neutral Point of View policy that requires editors to represent ‘fairly, proportionately, and, as far as possible, without editorial bias, all of the significant views that have been published by reliable sources on a topic’ (‘Wikipedia:Neutral point of view’ 2019). How far this is adhered to inevitably varies, with parallel articles articulating different knowledge and national points of view on the same topic.

The research involves analysing carefully selected multilingual sets of article about contemporary news stories relating to Russia. Two have been investigated in detail to date, with two further sets provisionally planned. In the research the following kinds of question have been considered: how much translation does there appear to be in Wikipedia? how easily can its presence be detected? what are the commonest sources for translated material? where is translated material located within multilingual article sets? how does translated material enter an article and what happens to it in the process? how does it evolve over time? and how and to what extent does translation contribute to knowledge and POV?

Working methods have combined the collection of quantitative data and detailed qualitative work on specific translation fragments with the aid of the following tools:

- Google Translate
- Page View Statistics
- Revision History Statistics
- Revision History Search (WikiBlame)
- WikiWho

¹The authors would like to thank the Hong Kong University Grants Council for funding this project (under the General Research Fund, project no. 12603619 ‘Understanding Wikipedia’s dark matter: translation, knowledge and point of view’).
Tools for edit timeline visualisation, tracking words in different language editions and detecting candidate translation pairs are currently under development and others may also follow if time permits.

To date, a number of findings have emerged from this analysis. While there is evidence of substantial translation activity, most individual translation acts are rarely apparent to readers. Some, but not all, are accompanied by a reference to a source in another language, which means that knowledge provenance is not always clear. The English and significant local editions are used almost exclusively as sources for translation between editions and for access to external sources in other languages. Three major scenarios describing how the presence of translated fragments develops over time (gradual similarity decay caused by changes to the target and/or source text, the appearance and subsequent disappearance of translated material, and its long-term persistence) have been identified along with at least six transfer types (e.g. whole article translation, translation of short fragments and the use of un-post-edited machine translation output). The Wikipedia principle of co-creation, moreover, appears to work across language editions. In terms of POV, translated material is generally either relatively neutral or supportive of an article’s own standpoint and can be used to introduce wording, quotes and more major article features either at the point of translation or incrementally thereafter. It therefore plays an important but generally unnoticed role in constructing an article’s POV.

Our current aims are to continue developing a robust and sophisticated methodology to facilitate the study of Wikipedia translation and to perform further detailed multilingual case studies with the aim of generalising and extending – or possibly challenging – current conclusions regarding the location, nature and development of translated material in Wikipedia as well as its contribution to POV and knowledge production. It will continue to examine POV across thematically linked articles within a single Wikipedia edition and will also start to examine the interaction between collaborativity and individual editors’ self-motivated action. Further research tools are likely to be developed as part of the project for various purposes. A blog documenting the project findings will also be produced.

References