
Machine Translation and Post-Editing for User Generated Content: An LSP Perspective

Elaine O'Curran

elaine.ocurran@welocalize.com

Welocalize Inc., 6 Dundee Park, Andover, Massachusetts 01810, USA

Abstract

User Generated Content (UGC) is a new and exciting content type for Language Service providers (LSP) and it poses its own distinctive challenges for a machine translation workflow: UGC requires more pre-editing steps than any other content type we process with MT and it demands non-traditional approaches to post-editing, resourcing and quality evaluations.

We discuss the most common quality level requirements that we have observed in our work with UGC and ways to achieve them using specific post-editing methodologies for this content type. We will also touch on the subjects of resource selection, our experience around MT engine evaluations and customization for this content type and the importance of using the appropriate evaluation method for different use cases.

1. Introduction

With social media content - such as blogs, travel reviews, online market places, and technical user forums - taking a very prominent place in companies' global marketing outreach, the need to provide this content globally is growing exponentially. For simple cost reasons, human translation is often not a viable option and the use of raw MT to publish this content is now a common approach in order to meet the demands of high volumes and high perishability. However, raw MT is not always delivering to the desired quality standards. Additionally, Google does not index content that is identified as machine translated, and as a consequence machine translated content cannot be found in Google searches. This is where post-editing to "just the right quality level" comes into play.

2. How useful is MT for UGC?

There are a number of challenges for MT due to the characteristics of this content type. A lot of UGC is authored by ordinary users who are not technical writers, marketing or media professionals and often may not even be native speakers of the language they are writing in. The style tends to be very informal and spoken in character, with spelling and grammar errors, and the use of non-standard input such as emoticons are commonplace. Add to that the huge multitude of authors, each with their own style and jargon, and we are left with an enormous lexical and stylistic diversity that cannot be found in traditionally authored content. (Roturier and Bensadoun, 2011)

Researchers are focusing efforts on normalization and preprocessing steps of UGC in order to improve MT output.

2.1. Evaluation Methods

Utility scoring can be used to measure the quality of raw MT, especially for the purpose of gisting. It is a commonly used human evaluation criteria to rate how understandable and usable the raw MT output is on a scale of 1 to 5 (see Figure 1 below). A score of 3 or higher is usually considered a pass and means the content is somewhat comprehensible and actionable, and hence publishable.

Score	Description
5	The document is understandable and actionable. Nearly all of the text is well translated. That is to say that you perfectly understood the document context, such as comprehending a property description, or a travel review.
4	The document is understandable and actionable. Most of the text is well translated. That is to say that you properly understood the document context, such as comprehending a property description, or a travel review.
3	The document is not entirely understandable, but it is actionable. The text is stylistically and grammatically odd. Some of the text is well translated. That is to say that the text contains many errors but you are still able to extract from it basic context, such as comprehending essential aspects of a property description, or a travel review.
2	The document is possibly understandable and actionable given enough context and/or time to work it out. That is to say that the text contains many errors and it is difficult to extract from it basic context, but given a lot of time, it could be deciphered to comprehend aspects of a property description, or a travel review.
1	The document is not understandable and it is impossible to understand the information it contains.

Figure 1: Utility Scoring Definitions

2.2. Results

Two studies of raw MT evaluation results for a technical user forum posts showed that roughly around 50% of the posts were comprehensible (Roturier & Bensadoun, 2011; Mitchell & Roturier, 2012).

We have performed a number of MT evaluations on travel content translated by a commercial third party engine that has been customized for the content and includes a number of normalization steps of the source text. We found that between 54% and 96% of reviews scored between 3 and 5 on the Utility scale (see Figure 2 below).

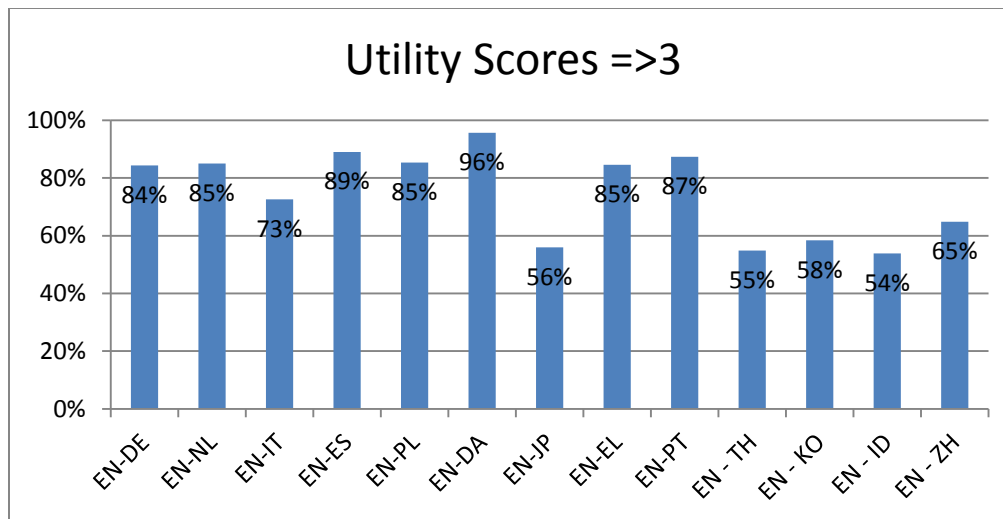


Figure 2: Utility Scores from 3 - 5.

3. Localization strategies for UGC

The bulk of UGC is published with raw MT, but each customer determines the impact that the UGC has on the business and brand, and this drives the content selection for post-editing. We generally post-edit the UGC content which is expected to deliver information – forums, reviews, knowledge bases - and only the *highest visibility* part of this content, or content that meets certain criteria, such as a high number of visits or clicks.

3.1. Light Post-Editing

The emphasis of this quality level is to manage quick turnarounds and large volumes. Here we see a sliding scale depending on the light PE strategy; this can be simply a sanity check to ensure that UGC content is not published with severe misrepresentations or offensive statements. Basically, any post-editing task that falls below the quality expectations of medium post-editing belongs into this category.

3.2. Medium Post-Editing

This quality level is close to human-level, but with style and fluency allowances. The emphasis is on meaning and readability, an interim between full and light. A good example of this level is technical forum content, where the aim is to provide technically accurate translations that will enable readers to solve the problem they are experiencing. Style and fluency are not important.

3.3. Full Post-Editing

Here we post-edit to human level, ensuring correct grammar, fluency, terminology, style & voice. Crowdsourcing is a viable option to reproduce the multitude of styles & voices found in UGC. For some content, it is important to allow the author's unique voice and personality to shine through and to avoid using a corporate tone. The aim is to maintain the emotional core of the message.

3.4. Transcreation

There are exceptional scenarios where UGC qualifies as high-end, high-visibility content and where a translation approach for marcom / transcreation is required. If the objective of the UGC content is to trigger *emotional impact* it is best not to post-edit such content at all and go with human translation. Examples of such content is usually *new marketing* - CEO blogs, first page product reviews, tweets - anything that's expected to convey the company image.

3.5. MT Corpus Management

This is a highly customized post-editing task that creates corpus data for MT engine training of user generated listings. The strategy is driven by client requirements and corpus purpose, but it is generally closer to full or medium post-editing.

4. Resourcing Strategies for Translating UGC

UGC translation is a task that can be well-suited to crowdsourcing. The chunks of content in a review or a thread are perfectly sized for translation or post-editing by a crowd. The usual concerns we have in traditional projects about limiting the size of the translation team to ensure linguistic and terminology consistency across a project do not apply to most UGC content. Here we value the lexical and stylistic diversity we can achieve through crowdsourcing that would be difficult to attain with a traditional resourcing model.

The ideal target groups are language graduates or bilingual people that are computer literate, internet savvy, competent at writing in their own language and have experience with the domain, i.e. travel, retail, technical products etc.

- Recent graduates looking for translation experience to start their careers
- Qualified linguists/bilinguals working in other areas (teaching, call centre support, etc.) who do not have professional translation experience or specialist domain expertise, but have the skills to engage in translation in a casual way
- Retired translators who want to do some light part-time work, maybe out of touch with the industry but still capable of translating
- Bilingual teens, undergrads with sufficient linguistic skills to produce "good enough" translation in social media/community context where information is targeted at youth

However, the crowd needs to be carefully vetted for their linguistic competency, and in some cases they must be able to learn to use a simple CAT tool as well as understand and implement simple translation or post-editing instructions.

5. Quality Assurance Models

We recommend a simple workflow for UGC in order to reduce the touch points that increase effort and reduce productivity. There is no traditional LQA or quality program required for light or medium post-editing or crowd translation. For these quality levels, there are a number of approaches that can be used:

- Automated spell-checking pass
- Basic xBench pass (terminology, source/target consistency)

- Community or end user feedback
- Level of quality measured by reach (unique visits, clicks, etc.)
- Readability scoring on a 4-point scale (only spotchecks) using TAUS Dynamic Quality Framework definitions.

It is important to add, however, that quality expectations will still need to be defined before project start on a per-client and per-content basis to make sure expectations are clear for all stakeholders.

6. Conclusions

When translating UGC and social content, we have to identify the appropriate quality levels and use tools like machine translation and crowdsourcing to address time and cost issues. There are a number of available tools and strategies to smoothly integrate the translation of UGC and social media content into the overall localization workflow. We can also expect the quality of the raw MT output to increase, as more time and research is invested in normalization processes for UGC.

Acknowledgement

The author would like to acknowledge the support of Olga Beregovaya, Laura Casanellas and Lena Marg.

References

- Mitchell, L. and Roturier, J. (2012). Evaluation of Machine-Translated User Generated Content: A pilot study based on User Ratings. In Proceedings of EAMT 2012, Trento, Italy, pp 61–64.
- Roturier, J. and Bensadoun, A. (2011). Evaluation of MT Systems to Translate User Generated Content. In Proceedings of Machine Translation Summit XIII, Xiamen, China, pp. 244–251.
- Jiang, J., Way, A, Haque, R. (2012). Translating User-Generated Content in the Social Networking Space. In Proceedings of AMTA2012, San Diego, California, USA, 9pp.