Unbabel

Quality at Unbabel

Unbabel ensures quality by using a framework inspired by the Multidimensional Quality Metrics (MQM) framework. MQM are originally developed within the QTLaunchPad european project. QTLaunchPad was a European Commission-funded collaborative research initiative (2012-2014) dedicated to preparing the grounds for a new type of systematic research and cooperation between MT developers and language industries for overcoming quality barriers in machine and human translation using language technologies. It's an open system including a flexible list of issue types.

Unbabel MQM provides an adapted framework for describing and defining quality metrics used to assess the quality of translated texts and to identify specific issues in those texts. It provides a systematic framework to describe quality metrics based on the identification of textual features and an important input to improve our translation systems.
Our objective is to have all the LP's assessed (90% of the content of our pipeline) at a 95 MQM, perceived in the industry as the professional level.

Unbabel ensures the quality of its translation by two checkpoints:

- **Assessment of the quality of the work delivered (fine-grained issues Annotation)**: Weekly Checkpoint and Proactive Quality Audits are done to ensure a permanent control and feedback to the client and also to our Editors. The fine-grained error categories adopted follow the typology established by QTLaunchPad. This typology was, however, adapted having in mind Unbabel's requirements and the types of data annotated
    1) **Accuracy.** Accuracy addresses the relationship of the target text to the source text and can be assessed only by considering this relationship. Examples of accuracy errors include changes in intended meaning, addition and omission of content or any type of mistranslation. In sum, the target text does not accurately reflect the source text.
    2) **Fluency.** Fluency addresses the linguistic well-formedness of the text that can be assessed without regard to whether the text is a translation or not. Examples of fluency errors include grammar or spelling issues. In sum, these errors affect the reading and the comprehension of the text.
    3) **Style.** Style usually relates to the use of register (formal vs. informal vocabulary) and compliance with the clients' instructions and glossary.
- **Assessment of the quality of the editors (rating 1-5)**: on the onboarding of the editors and through regular evaluations by high-level issues detection.

1) **First Onboarding Step**: Newly onboarded Editors need to complete a set of tasks in order to prove that they're not bots and that they actually accurately edited the text. This filtering is automatic and if they pass this test, they go the the trainee editor step. If they don't pass, they don't become Unbabel Editors.
2) **Trainee Editor Evaluation**: In order to become paid and being able to work on real tasks, Unbabel Editors are asked to complete a set of training tasks. At this step, they are evaluated by humans (PRO translators) following a high-level issues grid. If the Editor doesn't pass, he/she is out of the Unbabel translation pipeline. If he/she passes, then, he/she has access to Unbabel translations inside our pipeline.
3) **Paid Editor Evaluation**: Paid Editors are evaluated following the criteria used for trainee editors, on a regular basis. If he/she receive a rating below the minimum one, he/she is demoted to trainee and as no longer access to Unbabel paid translations.

Drawing on the examples of leading Quality Assurance systems and tools, Unbabel MQM provides a comprehensive and extensible list of quality issue types that can be used in several Quality Assurance tasks.

- **Separation of different aspects of translation.** MQM supports analysis of source and target fluency to help Quality team identify the source of problems and tackle them. It supports assessment of "end user adequacy" (ratings of whether the translation meets its requirements). By separating these aspects, Quality Team gain insight into the source of issues and how to resolve them.
- **Multiple levels of granularity.** Issues are categorized in a hierarchy, allowing our annotators to use high-level issue types for quick assessments during the evaluation checkpoints or fine-grained categories to support issue diagnosis and resolution.
- **Extensibility.** It's possible to add additional, or more fine-grained, issue types, there is an easy mechanism to do so. In addition, it is possible to adjust the quality measurement depending on client objectives (e.g. more weight on fluency than accuracy).