

Quality Evaluation using an Error Typology Approach

TAUS Signature Editions

Table of Contents

Why are TAUS industry guidelines needed?	4
Error Typology Best Practice Guidelines	5

Why are TAUS industry guidelines needed?

Error typology is the standard approach to quality evaluation currently. There is some consistency in its application across the industry, but there is also variability in categories, granularity, penalties and so on. It is a largely manual process, focused only on small samples and takes time and money to apply.

Providing guidelines for best practice will enable the industry to:

- Adopt a more standard approach to error typologies, ensuring a shared language and understanding between translation buyers, suppliers and evaluators
- Move towards increased automation of this quality evaluation mechanism
- Better track and compare performance across projects, languages and vendors

Error Typology Best Practice Guidelines

For quality evaluation based on the error typology, limit the number of error categories.

- The most commonly used categories are: Language, Terminology, Accuracy and Style.
- Diagnostic evaluations that seek to understand in detail the nature or cause of errors may require a more detailed error typology. For further details on error categories, refer to the [TAUS DQF Framework Knowledgebase](#). The Error Typology should be flexible enough to allow for additional or sub-categories, if required.

Establish clear definitions for each category.

- The commonly used category of 'Language' could be ambiguous, but an error in this category generally means a grammatical, syntactic or punctuation error.
- The category of 'Accuracy' is applied when incorrect meaning has been transferred or there has been an unacceptable omission or addition in the translated text.
- The category 'Terminology' is applied when a glossary or other standard terminology source has not been adhered to.
- The category of 'Style' can be quite subjective; Subjectivity can be reduced by defining this as 'Contravention of the style guide'. Where an error of this type occurs, reference should be made to a specific guideline within the target-language-specific style guide.
- List typical examples to help evaluators select the right category
- Add different weightings to each error type depending on the content type

Have no more than four severity levels.

- The established practice is to have four severity levels: Minor, Major, Critical and Neutral. 'Neutral' applies when a problem needs to be logged, but is not the fault of the translator, or to inform of a mistake

- that will be penalized if made in the future.
- Different thresholds exist for major, minor and critical errors. These should be flexible, depending on the content type, end-user profile and perishability of the content. For further information, TAUS DQF Framework Knowledgebase.

Include a positive category/positive action for excellent translations.

- Acknowledging excellence is important for ensuring continued high levels of quality. Translators often complain that they only receive feedback when it is negative and hear nothing when they do an excellent job.

Use a separate QE metric for DTP and UI text.

- Use a separate metric for these because specific issues arise for DTP (e.g. formatting, graphics) and for UI text (e.g. truncations).

Provide text in context to facilitate the best possible review process.

- Seeing the translated text as the end user will see it will better enable the evaluator to review the impact of errors.
- Allow reviewers to review chunks of coherent text, rather than isolated segments.
- Ideally, the translation should be carried out in a context-rich environment, especially if the quality evaluation is to be carried out in such an environment.

To ensure consistent quality human evaluators must meet minimum requirements.

- Provide training materials, screening tests, and guidelines with examples
- Evaluators should be native or near native speakers, familiar with the domain of the data
- Evaluators should ideally be available to perform one evaluation pass without interruption

Determine when your evaluations are suited for benchmarking, by making sure results are repeatable.

- Define tests and test sets for each model and determine minimal requirements for inter-rater agreements.
- Train and retain evaluator teams
- Establish scalable and repeatable processes by using tools and automated processes for data preparation, evaluation setup and analysis

Capture evaluation results automatically to enable comparisons across time, projects, vendors.

- Use color-coding for comparing performance over time, e.g. green for meeting or exceeding expectations, amber to signal a reduction in quality, red for problems that need addressing.

Implement a CAPA (Corrective Action Preventive Action) process.

- Best practice is for there to be a process in place to deal with quality issues - corrective action processes along with preventive action processes. Examples might include the provision of training or the improvement of terminology management processes.

Further resources

For information on when to use an error typology approach, detailed standard definitions of categories, examples of thresholds, a step-by-step process guide, ready to use template and guidance on training evaluators, please refer to the [TAUS Dynamic Quality Framework Knowledge Base](#).

Our thanks to:

Sharon O'Brien (TAUS Labs) for drafting these guidelines.

The following organizations for reviewing and refining the Guidelines at the TAUS Quality Evaluation Summit 15 March 2013, Dublin:

ABBYY Language Services, Capita Translation and Interpreting, CLS Communication, Crestec, EMC Corporation, Intel, Jensen Localization, Jonckers Translation & Engineering s.r.o., KantanMT, Lexcelera, Lingo24, Lionbridge, Logrus International, McAfee, Microsoft, Moravia, Palex Languages & Software, Safaba Translation Solutions, STP Nordic, Trinity College Dublin, University of Sheffield, Vistatec, Welocalize and Yamagata Europe.

Consultation and Publication

A public consultation was undertaken between 11 and 24 April 2013. The guidelines were published on 2 May 2013.

Feedback

To give feedback on improving the guidelines, please write to dqf@taus.net

About TAUS

TAUS is a think tank and resource center for the global translation industries. We develop communities through a program of events and online user groups and by sharing knowledge, metrics and data that help all stakeholders in the translation industry develop a better service.

The shared knowledge helps TAUS members decide on effective localization strategies. Shared metrics support more efficient processes and the normalization of quality evaluation. Shared data leads to improved translation automation.

TAUS develops APIs that give members access to services like DQF, the Quality Dashboard and the TAUS Data Cloud through their own translation platforms and tools. TAUS metrics and data are already built into most of the major translation technologies.