Who needs an MQM Scorecard?

Alan Melby & Catherine Marshall

TC43 | November 18, 2021

Multidimensional Quality Metrics (MQM)

http://themqm.org/ (the MQM website)

MQM is a framework for identifying and classifying errors in human or machine translation.

It uses an **analytic** translation quality evaluation (TQE) methodology. Not **holistic**.

It is a manual, reference-free approach. Thus, it allows human bilinguals to identify specific translation errors at the segment level. No reference translation is required, as with BLEU.

MQM assists **measuring** the quality of a translation using a custom metric, relative to a set of specifications and **inspiring** translation quality improvement.

MQM vs. Reference-based metrics

MQM is an example of a reference-**free** metric. There are also reference-**based** metrics such as BLEU or METEOR.

Reference-based metrics automatically compare the translation (whether human or machine) against one or more professional reference translations.

Reference-free metrics, such as MQM, don't require a reference translation. The translation is analytically evaluated by a human who understands the specifications of the project and thus the criteria that it is being evaluated on.

3

Related TQE Standards

Analytic Quality Evaluation (ASTM WK46396)

- **MQM (which was developed in EC projects) is the basis for this standard**
- WK46396 is expected to go to ballot in early 2022
- MQM has over 100 error types organized within seven overarching dimensions: Terminology, Accuracy, Linguistic Conventions (formerly Fluency), Style, Locale Conventions, Audience Appropriateness (formerly Verity), and Design & Markup

Evaluation of Translation Output (ISO 5060, under development)

Holistic Quality Evaluation (ASTM WK54884)

- Shaped by the material being perceived as a whole (on a macro level)

Needed to set up an MQM Scorecard project

MQM Typology (XML file)
- MQM's framework for defining translation quality metrics

Metric file (XML file)
- Subset of the Typology file based on the needs of the specific project

Bitext file (TXT file; can start from ordered TMX or XLIFF and converted)
- UTF-8 text file where each line is a single segment in the source text followed by a tab and then the corresponding target segment

Optional: Structured Specifications File (STS file)
- File containing the specifications relevant to the project (specifications are not optional but it is optional to load them into the Scorecard)

Current Projects using the MQM Scorecard

ATA Database Project (joint effort of KSU and BYU)

Whale Project (a project of FIT (www.fit-ift.org)

We are currently looking for collaborators to use the Scorecard for their projects.

If interested, contact Alan Melby

Running the MQM Scorecard

Browser based; nothing to install on client computer
- We have been using Firefox and Chrome

Data privacy; each Scorecard installation is self-contained

Behind the scenes
- 2015-2020 the Scorecard used a PHP framework (Symfony)
- In 2021, the Scorecard has been ported to a JavaScript framework
- Original version was a BYU MA project using typically tangled PHP
- Symfony version was developed in Germany by DFKI
- Port to the new framework has been funded by BYU and LTAC

BYOS: Bring your own server

Linux Server basic requirements:
- Root access to the server
- Can run React and Express

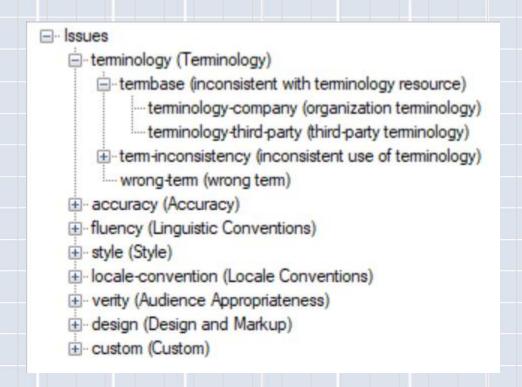Additional software requirements:
- Git (for a version control system)
- PostgreSQL v9.x
- Node.js v16.x

Minimum hardware Requirements:
- 1 GHz processor
- 2 GB of RAM
- 512 MB of HDD

DEMO

MQM Error Typology

☐ Issues
    ☐ terminology (Terminology)
        ☐ termbase (inconsistent with terminology resource)
            ---- terminology-company (organization terminology)
            ---- terminology-third-party (third-party terminology)
        ⊞ term-inconsistency (inconsistent use of terminology)
        ---- wrong-term (wrong term)
    ⊞ accuracy (Accuracy)
    ⊞ fluency (Linguistic Conventions)
    ⊞ style (Style)
    ⊞ locale-convention (Locale Conventions)
    ⊞ verity (Audience Appropriateness)
    ⊞ design (Design and Markup)
    ⊞ custom (Custom)

Creating a Project

MQM Evaluation Scorecard

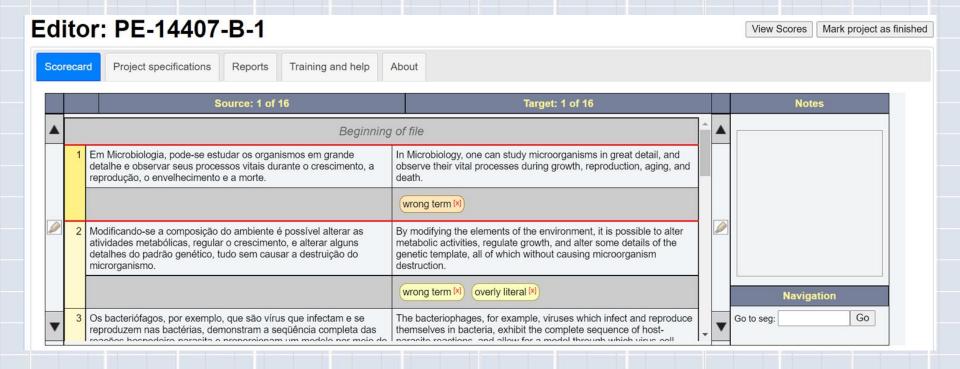Create project   View projects   Manage users   Mange Typology   Edit profile   Training and Help

Create Project

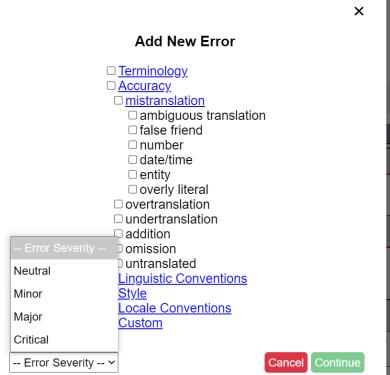| Project name	
Bi-text file (tab-delimited, UTF-8	Choose File No file chosen
Specifications file (optional)	Choose File No file chosen
Metric file	Choose File No file chosen
Submit	

Scorecard Interface

Adding a new Issue

Add New Error

- ☐ Terminology
- ☐ Accuracy
- ☐ Linguistic Conventions
- ☐ Style
- ☐ Locale Conventions
- ☐ Custom

-- Error Severity --

Cancel Continue

wrong term [x]   overly literal [x]

Add New Error

- ☐ Terminology
- ☐ Accuracy
  - ☐ mistranslation
    - ☐ ambiguous translation
    - ☐ false friend
    - ☐ number
    - ☐ date/time
    - ☐ entity
    - ☐ overly literal
  - ☐ overtranslation
  - ☐ undertranslation
  - ☐ addition
  - ☐ omission
  - ☐ untranslated
- Linguistic Conventions
- Style
- Locale Conventions
- Custom

-- Error Severity --
Neutral
Minor
Major
Critical

-- Error Severity --

Cancel Continue

Notes Section

Seeing Errors using the Interface

Error Summary

Issue	Source					Target					Total
	Neutral	Minor	Major	Critical	Subtotal	Neutral	Minor	Major	Critical	Subtotal	
Terminology											
Terminology	0	0	0	0	0	0	0	0	0	0	0
inconsistent use of terminology	0	0	0	0	0	0	0	0	0	0	0
multiple terms in translation	0	0	0	0	0	0	0	0	0	0	0
wrong term	0	0	0	0	0	0	1	1	0	2	2
Subtotal	0	0	0	0	0	0	1	1	0	2	2
Accuracy											
Accuracy	0	0	0	0	0	0	0	0	0	0	0
mistranslation	0	0	0	0	0	0	2	1	0	3	3
ambiguous translation	0	0	0	0	0	0	0	0	0	0	0
false friend	0	0	0	0	0	0	1	0	0	1	1
number	0	0	0	0	0	0	0	0	0	0	0
date/time	0	0	0	0	0	0	0	0	0	0	0
entity	0	0	0	0	0	0	0	0	0	0	0
overly literal	0	0	0	0	0	0	5	0	0	5	5
overtranslation	0	0	0	0	0	0	0	0	0	0	0
undertranslation	0	0	0	0	0	0	0	0	0	0	0
addition	0	0	0	0	0	0	0	0	0	0	0
omission	0	0	0	0	0	0	1	0	0	1	1
untranslated	0	0	0	0	0	0	0	0	0	0	0
Subtotal	0	0	0	0	0	0	9	1	0	10	10
Linguistic Conventions											

You can Export the Scorecard to a JSON file

Subtotal	0	0	0	0	0	0	0	0	0	0	0
Custom											
Custom	0	0	0	0	0	0	0	0	0	0	0
illegibility	0	0	0	0	0	0	0	0	0	0	0
indecision	0	0	0	0	0	0	0	0	0	0	0
ungraded	0	0	0	0	0	0	0	0	0	0	0
kudos	0	0	0	0	0	0	0	0	0	0	0
Subtotal	0	0	0	0	0	0	0	0	0	0	0
Total	0	0	0	0	0	0	22	3	0	25	25

Export Scorecard data (JSON)

17

JSON File Information

Project Name
Key (links)
Errors
        Segment

        Error-type Name
        Severity
        Notes
        Highlighting
            Start Index
            End Index
Score
Segments
        Source
        Target

Why might you need the MQM Scorecard?

To produce analytic, reference-free evaluations of translations. Via your own or a rented Linux server.

Note: There are more sophisticated MQM-compatible TQE tools available commercially.

Interested? contact alan.melby@fit-ift.org