

Translation Quality Standards

Dr. Arle Lommel

alommel@csa-research.com

March 17, 2018

Association for Machine Translation in the Americas

Boston, MA



This presentation covers*

Feature	WD ISO 21999	ISO 17100:2015	ASTM F2575-14	ASTM WK46396 (MQM-DQF)
Type and Status	Proposed committee draft for international standard (ballot closed)	International standard	Global standard (based on individuals and companies, rather than national bodies) — under revision.	Working draft based on industry standard from DFKI and TAUS
Guidance vs. requirements	Guidance	Requirements	Guidance, update will be requirements	Requirements (proposed)
Orientation (process vs. product)	Combination of process and product	Process (quality assurance)	Combination (primarily process)	Product
Includes quality metric?	Yes	No	No	Yes

- This presentation does not cover the following:
 - ASTM WK41374, for language service companies standard, which adopts a business process approach (in final editing)
 - ASTM WK54884, translation quality evaluation workflow (working draft)

Why do these standards matter for MT?

- Disconnect between MT development and commercial translation production that includes MT
 - Reference-based methods (such as BLEU et al.) is central to MT development
 - Reference-free quality evaluation is central to commercial translation production
 - See [MT Escaped from the Lab. Now What?](#) (Mike Dillinger, AMTA 2016)
- These standards are influential among *implementers* of MT
- Systematically improving MT requires knowledge about what MT does wrong and *why* one translation is better than another (requires analytic understanding)

What do we mean by translation
quality?

Well, what do we mean by “translation”?

- A translation (product) is target-language content corresponds to source-language content
 - Must include text
 - May include non-textual elements such as audio-visual content and software components
- Translation (process) is the action of creating target-language content that corresponds to source-language content according a set of specifications

What is translation quality?

- A subject of debate and contention
- Multiple measurement approaches are current
- Major disconnect between traditional human translation approaches and what MT developers do
- No one-size-fits-all set of specifications (more later)

Process vs. product

Process-centric	Product-centric
Emphasize the role of process steps on output quality	Focus on measuring conformance of translated content
Repeatability and consistency are the key	Can measure how well processes are delivering on expectations
Cannot determine whether the output meets requirements	Cannot verify process compliance or suitability

Universal or specifications-based

Universal	Specifications
There exists a set of one or more translations that are correct in all circumstances	Translations must conform to specifications: What works for one set may not work for another
Translations are good or bad	Translations are good or bad with respect to external factors
You could apply one set of criteria to any translation	Different requirements need different criteria

Reference-based or reference-free

Reference-based	Reference-free
Quality is measured by similarity to a known-good translation	Quality is measured by examining the product itself
Quick and cheap	Comparatively expensive and slow
Gives a score for the translation as a whole with no insight into specific issues	Provides insight into specific issues and allows decomposition (roll-down) of scores
Common in the MT world	Common in the HT world

WD ISO 21999: Translation quality assurance and assessment — Models and metrics

A general framework for translation quality

- Defines quality-related terms
- Sets up criteria for:
 - Job assignment and acceptance
 - Style
 - Consistency
 - Terminology
- Establishes a process for quality assessment of translation
- Discusses methods

Status

- Currently a committee draft
- Working through the ISO process
- Proposes a high-level typology of errors that is broadly compatible with MQM-DQF

ISO 17100:2015: Translation services – Requirements for translation services

A process-oriented approach to quality

- Defines processes that translation providers should follow to produce appropriate output
- Certifiable standard
- Replaces EN15038
- Widely implemented by language service providers
- Not generally applicable to MT

ASTM F2575-14: Standard Guide for Quality Assurance in Translation

Defines a stakeholder-oriented approach

- Emphasizes the role of specifications and factors that stakeholders need to consider
- Defines 21 parameters that need to be defined for projects
 - These serve as the basis for the MQM-DQF approach
 - Cover linguistic, production, environment, and relationship factors
- Apply to all translation projects, but does not define metrics for measuring quality
- The approaches in F2575 are largely ignored in MT circles, which results in a lack of clarity

ASTM WK46396: MQM-DQF (Translation Quality Assessment)

The F43 approach to quality

- ASTM F43 takes a *functionalist* approach, consistent with quality management principles:

A quality translation demonstrates *accuracy* and *fluency* required for the *audience and purpose* and *complies with all other specifications* negotiated between the requester and provider, taking into account both *requester goals* and *end-user needs*.

Genesis of WK46396

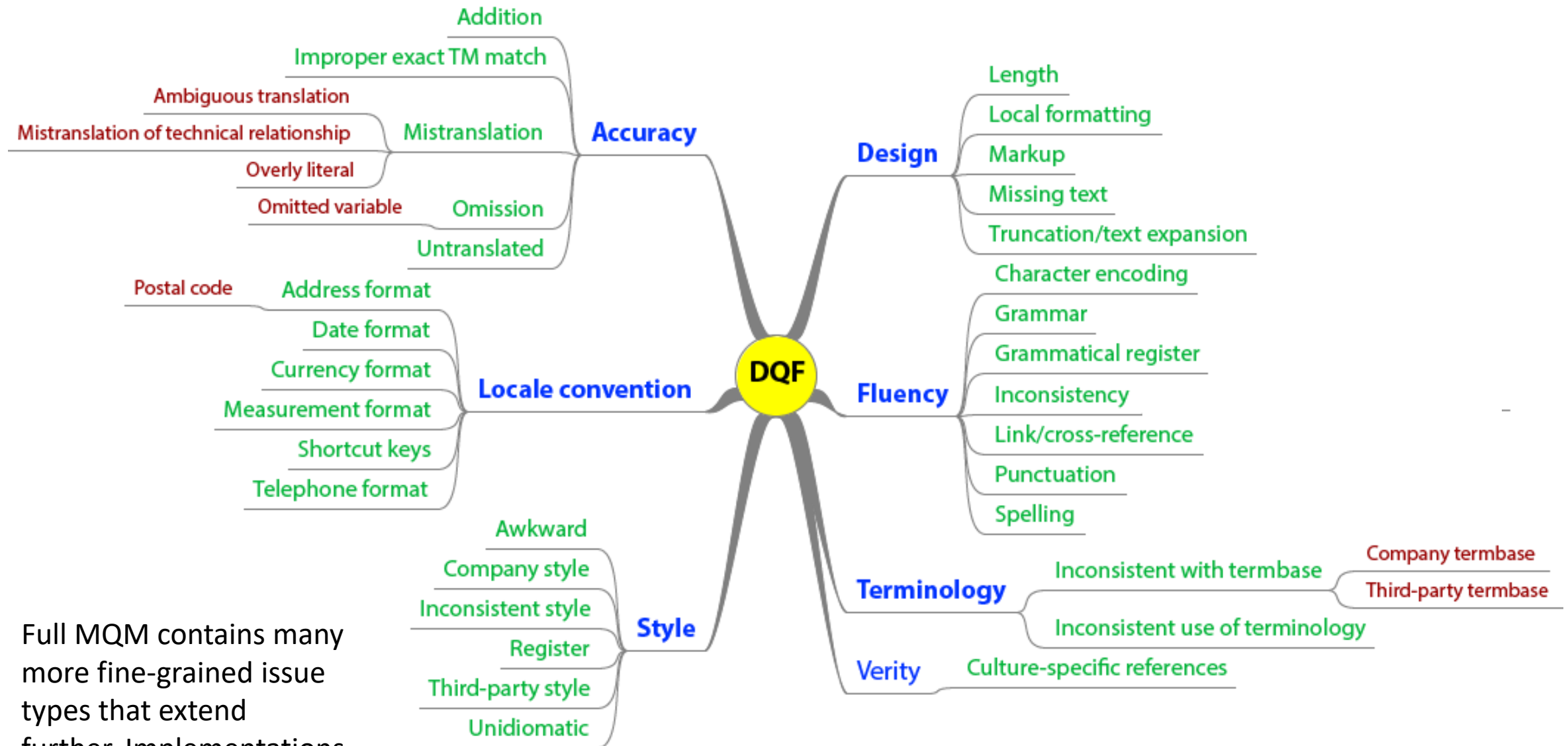
- TAUS Dynamic Quality Framework (DQF)
 - Arose from industry group
 - Focused on multiple approaches to quality (adequacy and fluency, error review, productivity measurement, MT ranking, and comparison)
 - Contained a relatively simple error typology
- DFKI Multidimensional Quality Metrics (MQM)
 - funded by European Commission in the QTLaunchPad and QTLeap projects
 - Focused on an exhaustive and robust *typology* of translation errors divided into multiple dimensions
- Unified error typology in 2014–2015 as a requirement of the European Commission

Drafting committee

- **Alan Melby** (BYU and LTAC Global)
- **Arle Lommel** (CSA Research)
- **Kirill Soloviev** (ContentQuo)
- **Aljoscha Burchardt** (DFKI)
- **Hans Uszkoreit** (DFKI)
- **Ingemar Strandvik** (European Commission)
- **Jennifer DeCamp** (MITER Corp)
- **Susanne Hempel** (SAP)
- **David Koot** (TAUS)
- **Jaap van der Meer** (TAUS)
- **Merle Tenney** (independent)

Approaches to quality evaluation

- Emphasizes analytic quality (identification and quantification of errors)
- Applies to both machine and human translation: Complements reference-based approaches such as BLEU
- Strives for consistency with ISO 9000-series approaches to quality management
- Emphasizes fairness
- Optionally considers source text as well as target text



Full MQM contains many more fine-grained issue types that extend further. Implementations use subsets

Focus of ASTM effort

- How to create custom metrics based on MQM-DQF
- Tied to translation specifications as defined in ASTM F2575
- Currently working on a complete draft to be ready by October
- Error typology will be maintained in W3C Community Group to ensure free and open access
 - Required by IP donors and European Commission (MQM was taxpayer funded)
 - W3C connection will promote closer ties with other standardization efforts and increase visibility for the error typology

Adoption of MQM-DQF

• **LSPs and Enterprises**

- Dell-EMC
- eBay
- LDS Church
- Lionbridge
- Microsoft
- Moravia
- Mozilla
- Seprotect
- Synergium
- Tableau
- Welocalize

• **Tools**

- ContentQuo
- MemSource
- SDL (plug-in)
- XTM

• **European projects**

- QTLaunchpad
- QT21

• **Academia**

- Various projects

• **In process**

- Argos Translations
- Booking.com
- CA Technologies
- Capita
- Crestec
- DaimlerIntuit
- John Deere
- Nike
- TNT-Fedex

Conclusion

Observations

- Translation quality is a hot topic
- A lot of people are trying to solve it
- The market is shifting towards MQM-DQF:
 - A good sign for the ASTM WK46349 effort
- ISO 17100 has high recognition among LSPs
- ISO 21999 is still in its infancy
- We still need to do more to bridge the gap between the BLEU/METEOR-centric approach of MT developers and the needs of consumers, translators, and LSPs