Multidimensional Quality Metrics: A New Unified Paradigm for Human and Machine Translation Quality Assessment

Hans Uszkoreit, Arle Lommel
DFKI (German Research Center for Artificial Intelligence)

Funded by the 7th Framework Programme of the European Commission through the contract 296347.
Outline

• The project QT Launch Pad
• The reasons for a metric shared by HT and MT
• The notion “translation quality”
• The reasons for a shared metric
• The basic concepts and components of the metric
• Insights obtained from the Roundtable/Tutorial
• Next steps
Consortium

- DFKI – Hans Uszkoreit (Coord.)
  Aljoscha Burchardt Project Manager, Stephan Busemann Administr. Coord.
- CNGL DCU – Josef van Genabith
- U. Sheffield – Lucia Specia
- ILSP Athena – Stelios Piperidis

- Subcontractor: GALA
  Hans Fenstermacher, CEO and former President
  Laura Brandon, GALA Managing Director
  Serge Gladkoff, GALA Standards Director
Main Objectives

• The support action will prepare the grounds for a new type of collaborative MT research dedicated to overcoming existing quality barriers.

• To this end, QTLaunchPad will...

  ▪ assemble and provide needed data and tools including specialised translation corpora, test suites and tools for quality assessment,
  ▪ create a shared quality metrics for human and machine translation, improve automatic translation quality estimation,
  ▪ extend an existing platform for resource-sharing to the needs of quality MT research,
  ▪ define strategies and challenges and then plan and launch a large-scale research and innovation action for a breakthrough in quality translation technology.
“The large MT action” could be one big project or, more likely, a cluster of several cooperating projects.

It may have a lead project and several satellite projects.

Satellite projects could be language oriented, e.g. Czech Satellite, or they could be research-theme oriented, e.g., Parsing Satellite, or they also be application task specific e.g. Pharma Satellite, Subtitling Satellite.

It should start within the duration of QTLP and serve as a trail blazer for a European push toward overcoming language barriers in HORIZON 2020 and maybe also CEF.

One scenario: A lead project funded by FP 7 and Satellites funded by national programmes, maybe also by regions or industry
is the working name of the grand endeavor, standing for

Quality Translation Technology for the 21st Century

It should serve as a trail blazer for a European push toward overcoming language barriers in HORIZON 2020 and maybe also CEF.

One scenario: A lead project funded by FP 7 and Satellites funded by national programmes, maybe also by regions or industry
The big Picture: Three Steps

- QTLaunchPad
- “The large MT action”
- CEF MT Service Deployment
- H20 Cluster Translation Cloud
Research Approaches

• highly specialized quality services,
• an analytical systematic fight against quality obstacles,
• improved utilization of shared data and tools,
• the integration of translation providers in research, deployment, testing and business models,
• and a concerted competitive model of technology evolution.
A New Approach

We view the planned new mode and approach of research as a paradigm shift. Future emerging research paradigms cannot be safely predicted but the envisaged mode of research should increase the chances for the development of completely new methods.
We view the planned new mode and approach of research as a paradigm shift. Future emerging research paradigms cannot be safely predicted but the envisaged mode of research should increase the chances for the development of completely new methods.
A New Approach

- Recognize the truly good
  High Quality Estimation

- Help the translator to improve the bad:
  Computer Assisted Translation

- Make the nearly good
  truly good:
  High Quality MT

- Help the end user to make the ugly at least understandable:
  Improve Gist Translation
Participating in the Process:

- Jan Hajic of U. Prague,
- Stephan Oepen of U. Oslo,
- Philipp Koehn of U. Edinburgh,
- Alex Waibel of Karlsruhe KIT,
- Marcello Federico of FBK Trento,
- Mikel Forcada of U. Alicante,
- Hermann Ney and Volker Steinbiss of RWTH Aachen,
- Nuria Bel of U. Barcelona,
- Joseph Mariani of LIMSI Paris,
- Johann Roturier of Symantec Ireland,
- Spyridon Pilos of EC DGT Luxembourg,
- Serge Gladkoff, Kim Harris and Hans Fenstermacher of GALA.
- Andrejs Vasiljevs, Tilde
Problem: No Shared Quality Metric

- Quality measured by BLEU, NIST, METEOR etc. does not indicate the type of quality problems
- These metrics are also better suited for measuring progress in the “ugly” and “bad” sectors of the quality spectrum
- Even the human evaluations usually by ranking, often done by CS researchers and students, do not help the human translators
- The LISA QA Model, EN-15038 and current ISO work on a successor, are not known and not used in MT research, neither ISO/TS-11669 the 20 criteria for translation projects
Current QA methods confuse distinct things:

- **Product Quality**
  - Fluency: *How well does the text read, independent of how it was produced?*
  - Accuracy (or adequacy): *Does the target say the same thing as the source?*
  - End-user or purpose adequacy: Does the text (source or target) meet its communicative purposes and does it reflect the real-world conditions around it?

- **Process**: *How is the translation produced? effectiveness, QA*

- **Project**: *How satisfactory is the project overall? processes, times, costs, results*
Cooperations

• Cooperation with the Standards Committee of GALA
• Cooperation with the Standards Committee of FIT (Fédération Internationale des Traducteurs / International Federation of Translators)
• Consultation with large client companies such as SAP, Paypal, Volkswagen, etc

• Hopefully in the future also cooperation with TAUS
What Is Translation Quality?

• A quality translation (1) demonstrates required accuracy and fluency (2) for the audience and purpose and (3) complies with all other negotiated specifications, taking into account end-user needs.

–Alan Melby
On quality

• Quality is relative.
  - It is an ideal target that depends on several given factors such as purpose, audience and budget. The target can be explicitly specified before translation. A useful quality metric must be suited for such quality specification.

• Quality is multidimensional, composed of several components that each may be measured by several criteria.
  - There are three types of criteria: the quality of the resulting text (textual quality, fluency), the quality of the translation (felicity, accuracy, adequacy), and formal quality (formatting, tags, fonts, etc. often also subsumed under fluency).

◆ Quality is also gradient in nature. Most criteria are scalar, they can be approximated.
  - The specified target usually is some composite approximation of an ideal target.
  - For many criteria, the distance to the ideal of minimal target can be measured by the frequency and severity of errors (error density).
  - In addition to a composite target there may minimum targets for the individual criteria.
Goal: Simplicity + Sophistication

- For different tasks in research and practical QA, different levels of sophistication are required. Sophistication translates into complexity.
- Nobody wants to deal with unneeded complexity.
- On the other hand, any standard that does not accommodate the degree of sophistication needed for research and for complex diagnostic assessment will be dismissed by some of the most relevant players.
- Thus our philosophy: Have a complex model, of which everyone can use just the parts required for some actual purpose.
Translation today is beyond human scale.

– Paula Shannon (Lionbridge)
MQM Principles
Background: QA Today

- QA often ad hoc (does the reviewer like it or not?) or even not done, based on trust
- MT and HT treated as different things
- The target is examined...
  - ...but the source is not (translators = garbage collectors?)
- Some standard metrics (SAE J2450, LISA QA) used as one-size-fits-all solutions
- Difficult to fix things systematically or identify the cause of problems
Translators are the garbage collectors of the documentation world.

– Alison Toon (HP)
I. Objectivity

- MQM provides a catalog of issue types with definitions and examples (120+)
  - We know that is too many
  - You are expected to use only a small fraction, as needed
  - How can I know *which* to use? (We’ll get to that)
- MQM provides a way to exchange information about the metrics you use
- Provides compatibility with new ITS 2.0 quality markup (you can use an ITS 2.0 profile for full compatibility)
- MQM provides an inline tagging format for XML that allows you to unambiguously specify what issues you see
- Goal is to move away from subjective feelings to objective issues
A Note: Issues vs. Errors

• **Issues** are potential problems with a text. (E.g., a non-matching price in a translated ad)
• **Errors** are actual, verified problems with a text.
  o If the non-matching price is correct (due to market differences) it is not an error
  o If the non-matching price is incorrect (due to mistranslation) it is an error
• MQM defines issue types. Only a human can determine if they are errors
2. No “Magic Bullet”

• A legal treaty ≠ a service manual ≠ a “gist” translation ≠ a summary translation
• If you assess with the wrong metric, you waste time and money
• Assessment should determine fitness for purpose
• MQM relies on the notion of dimensions (questions about the text: e.g., what domain? what language? what purpose?) to select appropriate issues
• By using metrics (ISO-TS/11669) from the beginning, you can develop a metric to meet needs from the start
• (Tyler will show a demo of working with dimensions to create a metric.)
3. Shared Systems

• While metrics may differ, they should share a common framework so users can tell how they are similar or different
• Metrics should be declarable in a shared vocabulary so they can be shared
• We intend different (vertical) industry groups to define their own shared metrics. E.g., a group of pharma companies could define a pharma metric in MQM
• QTLaunchPad is developing open-source tools for working with MQM that anyone can use
4. Consider the Source

- MQM (optionally) allows you to evaluate the source
- If source evaluation is included and the translator corrects problems in the source (i.e., delivers a translation that is “better” than the source), quality scores over 100% are possible!
- Translators deserve credit for fixing problems
- Translators need a way to mark issues in the source so they can be fixed once and for all (no more “didn’t I fix this the last twenty times already?”)
5. Good for Human and Machine Translation

- MT has been evaluated by (in essence) edit distance or by similarity to a reference corpus.
- MT scores are not comparable to human scores and tell us nothing about the *kinds* of errors.
- We cannot compare translation methods if we can’t evaluate them in a similar fashion.
- MQM makes a break from traditional MT evaluation methods (but allows them to be used as *additional* data).
- (We aren’t alone in this approach: in the last few months MT evaluation tools that use human-type metrics have been springing up.)
6. Free and Open

- Methods and tools need to be free and open
- MQM does not charge for any of its deliverables
- Final results will be released under an open-source license at least as permissive as the Eclipse Public License:
  - Developers can put MQM in commercial products, use it for exchange, etc.
- Developers can extend MQM as needed (add categories)
- Goal is to submit MQM for consideration as a standard framework for describing quality metrics
Introduction to MQM Issues
Types and Dimensions

Arle Lommel
**MQM Consists of...**

- A hierarchical catalog of issue types
  - Support various levels of granularity
  - Divided into three four branches:
    - Accuracy
    - Fluency
    - Verity*
    - Design
  - Features a core and extensions
    - Core supports coarse granularity and common issues
    - Extensions support additional needs and more granularity
- A set of twelve “dimensions” (based on ISO/TS-11669) to guide users in selecting appropriate issue types
- A method for declaring/describing a particular metric
MQM Consists of... (2)

• An inline format for tagging issues in XML files (currently being revised)
• A reporting format with scoring formula for determining scores/acceptance
MQM Based on Examination of...

- Manual
  - LISA QA Model
    - UI
    - Doc level 1
    - Doc level 2
  - SAE J2450
  - ISO 14080 (withdrawn)
  - SDL TMS Classic
  - ATA
  - CASL LPET
- Automatic
  - ApSIC XBench
  - CheckMate
  - XLIFF:doc
  - QA Distiller
  - Acrocheck
  - LanguageTool

- MT
  - BLEU
  - NIST
  - METEOR
  - WER/PER/(h)TER/TERp
  - Post-editing effort
  - Addicter
  - AMEANA
  - BLAST
  - DELiC4MT
  - Hjerson
  - TerrorCat
  - Woodpecker
What Is Translation Quality?

• A quality translation (1) demonstrates required accuracy and fluency (2) for the audience and purpose and (3) complies with all other negotiated specifications, taking into account end-user needs.
  –Alan Melby
Different Aspects of Quality

• Accuracy: Does the target say what the source says?
• Fluency: How well does the text say what it is supposed to?
• Verity: Does the text (source or target) say what it is supposed to?
• Design: Does the text look like it should?
Issue Types: All

MASSIVE OVERKILL!
(But somebody needs each of these)
Issue Types: Higher Granularity

- Accuracy
  - Bilingual terminology
  - Mistranslation
  - Omission
  - Untranslated
  - Addition

- Design
  - Overall design (layout)
    - Local formatting
    - Markup
    - Whitespace
    - Graphics and tables
    - Truncation/text expansion
    - Length
    - Internationalization

- Verity
  - Compleness
  - Legal requirements
  - Locale applicability

- Other

- Issue Types
  - Monolingual terminology
  - Inconsistency
  - Register
  - Syntactic
  - Duplicate
  - Spelling
  - Orthography
  - Grammar
  - Locale violation
  - Nonallowed characters
  - Nonallowed pattern
  - Nonconformance
  - Broken link/cross-reference
  - Unintelligible
  - Index/TOC

A LITTLE LESS
MASSIVE OVERKILL!
Issue Types: The Core

Accuracy
- Bilingual terminology
- Mistranslation
- Omission
- Untranslated
- Addition
- Fidelity
- Comprehensiveness
- Verity
- Legal requirements
- Locale applicability

Issue Types

Fluency
- Monolingual terminology
- Inconsistency
- Register
- Style
- Spelling
- Orthography
- Grammar
- Locale applicability
- Unintelligible

Much Better!
A Custom Metric

MORE REALISTIC
Sample mapping (SAE J2450)
Sample Mapping (LSP metric)

<table>
<thead>
<tr>
<th>Words in sample:</th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major weighting (if not specified)</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Score card</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross references</td>
<td>4</td>
</tr>
<tr>
<td>Omission/Addition</td>
<td>1</td>
</tr>
<tr>
<td>Incorrect Translation/Measuring</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Language</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Punctuation</td>
<td>4</td>
</tr>
<tr>
<td>Spelling/Typo</td>
<td>1</td>
</tr>
<tr>
<td>Grammar/Syntax</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Terminology</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Context</td>
<td></td>
</tr>
<tr>
<td>Inconsistency</td>
<td></td>
</tr>
<tr>
<td>Glossary</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Style</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>General Style</td>
<td></td>
</tr>
<tr>
<td>Client</td>
<td>Style Guide</td>
</tr>
<tr>
<td>Language Variants/Slang</td>
<td></td>
</tr>
<tr>
<td>Register/Tone</td>
<td></td>
</tr>
<tr>
<td>Necessity Additions</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Country</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Country/Rational Standards</td>
<td></td>
</tr>
<tr>
<td>Local Market Suitability</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Functional</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Format</td>
<td></td>
</tr>
<tr>
<td>Hidden Text</td>
<td></td>
</tr>
<tr>
<td>Tips/Unics</td>
<td></td>
</tr>
<tr>
<td>Technical Procedures</td>
<td></td>
</tr>
<tr>
<td>Spacing</td>
<td></td>
</tr>
</tbody>
</table>

| TOTAL | 14 | 18 |

<table>
<thead>
<tr>
<th>MQM scorecard</th>
<th>Minor</th>
<th>Major</th>
<th>Subsection score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>Bilingual terminology</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Mismarriage</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>(Do not translate)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Omission</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Untranslated</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Addition</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Accuracy Subtotal</td>
<td>4</td>
<td>4</td>
<td>99.0%</td>
</tr>
<tr>
<td>Quantity</td>
<td>Content</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Monolingual Terminology</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Inconsistency</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Register</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Variants and/or slang</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Register/subtolal</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Style</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Style, company</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Style subtolal</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Content subtotal</td>
<td>0</td>
<td>0</td>
<td>100.0%</td>
</tr>
<tr>
<td></td>
<td>Mechanical</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spelling</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Orthography</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Grammar</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Locale violation</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Broken link/cross-references</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Mechanical subtotal</td>
<td>10</td>
<td>14</td>
<td>96.7%</td>
</tr>
<tr>
<td>Quantity Subtotal</td>
<td>10</td>
<td>14</td>
<td>96.7%</td>
</tr>
<tr>
<td>Design</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Local formatting</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Markup</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>WhiteSpace</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Design Subtotal</td>
<td>0</td>
<td>0</td>
<td>100.0%</td>
</tr>
<tr>
<td>Verify</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Locale applicability</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Verify Subtotal</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Verify subtotal</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TOTAL</td>
<td>14</td>
<td>18</td>
<td></td>
</tr>
</tbody>
</table>
# Sample Score Card (Excel)

<table>
<thead>
<tr>
<th>Issue Type</th>
<th>Weight (default = 1)</th>
<th>Issue Counts</th>
<th>Penalty</th>
<th>Target Subscore</th>
<th>Source Subscore</th>
<th>Adjusted (Delta) Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accuracy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bilingual Terminology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mistranslation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do not translate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Omission</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Untranslated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Addition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Accuracy subtotal</strong></td>
<td>0</td>
<td></td>
<td>100.0%</td>
<td></td>
<td></td>
<td>100.0%</td>
</tr>
<tr>
<td><strong>Fluency</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unintelligible</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Fluency subtotal</strong></td>
<td>0</td>
<td></td>
<td>100.0%</td>
<td></td>
<td></td>
<td>100.0%</td>
</tr>
<tr>
<td><strong>Verity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fidelity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Completeness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Legal requirements</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Locale applicability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Fluency subtotal</strong></td>
<td>0</td>
<td></td>
<td>100.0%</td>
<td></td>
<td></td>
<td>100.0%</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>0</td>
<td></td>
<td>100.0%</td>
<td></td>
<td></td>
<td>100.0%</td>
</tr>
</tbody>
</table>
translate5: Tagging issues in place

This file is a based on a part of the php-online-Documentation. It's translation is done by a pretranslation based on a very fast winalign-Project and is not at all state of the translation art. It's only purpose is the generation of demo-data for translate5.

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Source text</th>
<th>Target text reworked</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>This file is a based on a part of the php-online-Documentation. It's translation is done by a pretranslation based on a very fast winalign-Project and is not at all state of the translation art. It's only purpose is the generation of demo-data for translate5.</td>
<td>Diese Datei ist Teil der php-online-Dokumentation. Ihre Übersetzung ist durch eine Vorübersetzung entstanden, die auf einem sehr schnell durchgeführten winalign und in keiner Art und Weise der Art eines Übersetzungsprojekts entspricht. Sein einziger Zweck ist die Erzeugung von Demo-Daten für translate5.</td>
</tr>
<tr>
<td>2</td>
<td>Apache 2.x on Unix systems - Manual</td>
<td>Apache 2.0 auf Unixsystemen - Manual asd</td>
</tr>
<tr>
<td>3</td>
<td>PHP Manual</td>
<td>PHP Manual asd</td>
</tr>
<tr>
<td>4</td>
<td>Installation and Configuration</td>
<td>Installation und Konfiguration</td>
</tr>
</tbody>
</table>
The 12 Dimensions

1. Language/Locale (source/target)
2. Subject field/domain
3. Terminology (source/target)
4. Text type
5. Audience
6. Purpose
7. Register
8. Style
9. Content correspondence
10. Output Modality
11. File format
12. Production technology

(See http://www.ttt.org/specs)
• Defining dimensions tells you important information about the text
• A selection of issue types should determine whether or not the text meets the requirements set out in the dimensions
• If issue meet the dimensions and quality problems are found that are not covered by the issues:
  o a. The dimensions were not specified properly (but the translator cannot be held accountable for not meeting what was not required)
  o b. The problems aren’t actually problems (for the specific requirements)
• E.g., For a gist translation, grammar may not matter. If grammar is tagged in a gist translation, it should not be tagged. (If grammar errors are so bad that they change the meaning, then the issue is *Mistranslation*, not *Grammar*)
Issues are selected using a question-based approach

We need your feedback on these questions
  - We are not the experts on practical needs
  - We need domain experts to contribute
  - Will be opening a doc to collect information
How Will This Be Used?

- Users won’t see this complexity.
- (Most) users will select a metric from a list and simply use it.
- Tools should bake in support (already have some preliminary commitments)
- If you need a new metric, you will answer some questions and get a recommendation or you will use a recommended third-party metric
Issues from the Roundtorial

- Automation
- Ergonomics
- Metrics need to lead to incremental improvement, not just measuring
- Need for speed/integration with Agile
- Training
- Demonstration of validity through multiple channels
Next steps

- Testing (we need your help)
- Implementation in open source tools (translate5 and others)
- Standardization: the community needs to own this
  - Anticipated start of committee work: September
  - Aiming for additional implementations in early 2014
  - Moving to standards body when committee consensus is reached
- Specialization within the system:
  - Domain- and task-specific extensions
  - Engagement with vertical industry groups
- Engagement with ISO TC 37
Live doc

Contact us...

- **Email**: info@qt21.eu
- **Web**: http://www.qt21.eu/launchpad
- **Facebook**: https://www.facebook.com/qtlaunchpad
- **LinkedIn**: http://bit.ly/qtlp-on-li