Measuring Quality in Translation

The Translation Quality Index and Other Methods

44th ATA Conference Phoenix 2003

Riccardo Schiaffino – President – Aliquantum

Franco Zearo – Worldwide Director, Language Services – Lionbridge Technologies

Overview

- Why Is Quality Measurement Important?

- How to Set Up a Quality Measurement System

- Demo of a Translation Quality Measurement Tool Prototype

- Practical Recommendations

- Questions & Answers

QC v QA

- Quality Control (QC)
  - Quality verification over the **whole** text.

    Example: Editing

- Quality Assurance (QA)
  - Sampling techniques, control of quality over a (statistically significant) **sample** of the whole text.

    Appropriate use: Quality measurement

When we study translation quality, we can focus on different things:

The translator

The translation process (the "process")
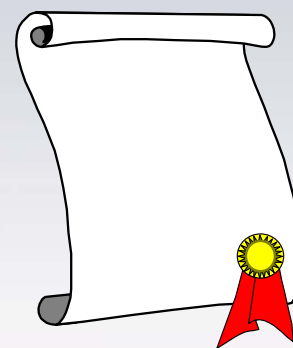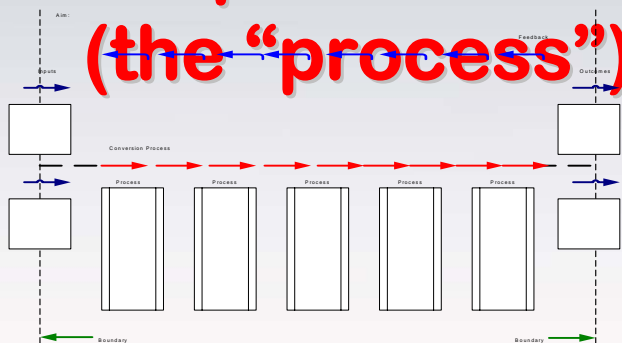
The translated text
(the "product")

Translation Quality Initiatives

ATA and other translators' certification initiatives

The translator

The translation process

DIN 2345
ISO 900x
UNI EN
10754
EUATC
ASTM

SAE J2450
LISA QA

The translated text

Very Important

- Improvements made to the overall process should result in improvements to the product (the translation)

- Measurements of the product quality should indicate if there have been actual improvements to the process

- Therefore, means to measure product quality must be in place

Why is Quality Measurement Important?

It is difficult to improve something if you cannot measure it

Metrics provide:

- A way to objectively quantify a process
- A means to reduce the cost of poor quality
- A means to increase customer satisfaction
- An opportunity for benchmarking
- Competitive advantages

Criteria for Successful Quality Measurements

Translation quality measurements should be:

– **Repeatable** (two assessments of the same sample yield *similar* results)

– **Reproducible** (different evaluators should arrive at a *similar* assessment for the same piece of translation

– **Objective** (void of subjective bias)

Measurement through Circumstantial Evidence

- Errors are circumstantial evidence of quality

- We believe that precise error measurement provides sufficient indication of good and bad translations

- A good translation is a translation with very few errors or none at all

How to Set Up a Quality Measurement System – Stage 1, Preparation

1. Collect examples of good and bad translations

2. Analyze the examples to separate controversial issues from agreed upon errors

3. Decide what to measure (error categorization)

4. Define what to measure in as many details as necessary (error definition)

How to Set Up a Quality Measurement System – Stage 2, Calibration

5. Assign a weight to various types of errors

6. Determine critical errors (if necessary)

7. Repeat 3, 4, 5, and 6 until the system works in an objective, repeatable, and reproducible way

Classification of Errors

An example of of an error classification process (not complete)

Start

Error? — No → End

Yes

Best detected by translators

May be detected by other users (including translators)

Meaning or form?

Meaning

Form

Mistranslation or Addition/ Omissions?

Grammar or Style?

Mistranslation

Addition/Omission

Style

Grammar

Misinterpretation of SL Miswriting of TL etc.

Addition Omission etc (?)

Usage or "Style Proper"?

Syntax Morphology Spelling Punctuation etc.

Usage

Style

Inconsistency Non-adherence to guidelines etc.

Register etc.

Definition of Errors

Deal with errors only when they violate agreed upon protocols of engagement whether implicit or explicit

Examples of explicit and implicit criteria:

- *Non-compliance errors (e.g. not following instructions)*
- *Violations of generally accepted language conventions*

Summary: Error Categorization

- **Select a (small) set of categories**

    - CTQ: Critical-To-Quality categories

- **Provide clear definitions**

- **Assign a weight**

    - Critical, Major, Minor

Why Do We Make Errors?

- The reasons behind the errors are separate from the measurement of the errors: Studying why errors happen is important, but it pertains more to quality control and improvement than to quality assurance

- E.g., capitalization errors due to the "Autocorrect" (mis)feature of MS Word (e.g., HBsAg "corrected" to HbsAg)

Correct Translation

A correct translation is a translation with no errors or where total error points result in a Translation Quality Index above the desired threshold

Quality Assurance Forms and Tools

Create a QA form (or a tool) to help graders give objective scores

Quality Assurance Form

Language:	Reviewer:	Date:	Result:	*Pass*	Comments:
Client Name					
Project Name					
Project Number					
Project Manager					
		Critical	max. error points + 1		
Number of words	0	Major	5 points		
Max error points allowed	0	Minor	1 point		

Error Category	Minor	Major	Critical	total	max. allowed
Mistranslation	0	0	0	0	0
Accuracy	0	0	0	0	0
Terminology	0	0	0	0	0
Language	0	0	0	0	0
Style	0	0	0	0	0
Country	0	0	0	0	0
Consistency	0	0	0	0	0
			Total	0	0

More elaborate descriptions of the error criteria can be found in the LISA QA model version 1.0 Reference Manual.

2 - Within the **Accuracy** category, give appropriate weight to the four following items (total must add up to 100%)

Accuracy		
	Incorrect meaning	40%
	Non-standard terminolgy	20%
	Inconsistent terminolgy	20%
	Untranslated SL	20%
	Total	100%

3 - Within the **Style** category, give appropriate weight to the three following items (total must add up to 100%)

Style		
	Wrong register	40%
	Inappropriate anglicisms	30%
	Inappropriate use of passive/active voice	30%
	Total	100%

4 - Within the **Grammar** category, give appropriate weight to the five following items (total must add up to 100%)

Grammar		
	Spelling errors	20%
	Typos	15%
	Grammar errors	35%
	Syntax errors	25%
	Punctuation errors	5%
	Total	100%

5 - Within the **Formatting** category, give appropriate weight to the five following items (total must add up to 100%)

Formatting		
	Layout errors	50%
	Font errors	40%
	Double spaces	10%
	Total	100%

Test1.doc - Microsoft Word

File Edit View Insert Format Tools Table Window Help Content Manager (M) Trados

Normal Times New Roman 12 **B** *I* U ... Error Log

How to Set Up a Quality Measurement System – Stage 3, Sampling

- Sampling
  - Selection criteria (e.g. random, systematic)
  - Size considerations (the greater the sample, the more accurate the results)
  - Select confidence intervals, margins of error
  - Cost considerations (find the point of diminishing returns)

How to Set Up a Quality Measurement System – Stage 4, Measurement

- Measurement

    – Evaluation must be repeatable, reproducible, objective

    – Use of independent auditors

    – Calculation of a Translation Quality Index (TQI)

Translation Quality Index (TQI)

The TQI is a number—obtained by the rigorous application of a QA process—that indicates the quality of a given translated text

Let's Calculate Two TQIs

| LISA QA Model ver. 1.0 (1995)	ATA Framework for Standard Error Marking
3,000 words (12 pages @ 250 words)	250 words (estimate)
30 error points	17 error points
30 error pts / 3,000 words = 0.01	17 error pts / 250 words = 0.068
Implicit TQI = 99.0%	Implicit TQI = 93.2%

Delusions of Accuracy

"Averages can be calculated to nineteen places of decimal with astonishing ease.

When the job is done, it looks very accurate.

It is an easy and fatal step to think that the accuracy of our arithmetic is equivalent to the accuracy of our knowledge about the problem in hand."

M.J. Moroney, *Facts from Figures*

How to Set Up a Quality Measurement System – Stage 5, Statistical Analysis

Investigate the Outcome

- At this stage there shouldn't be any special causes (use of control charts)
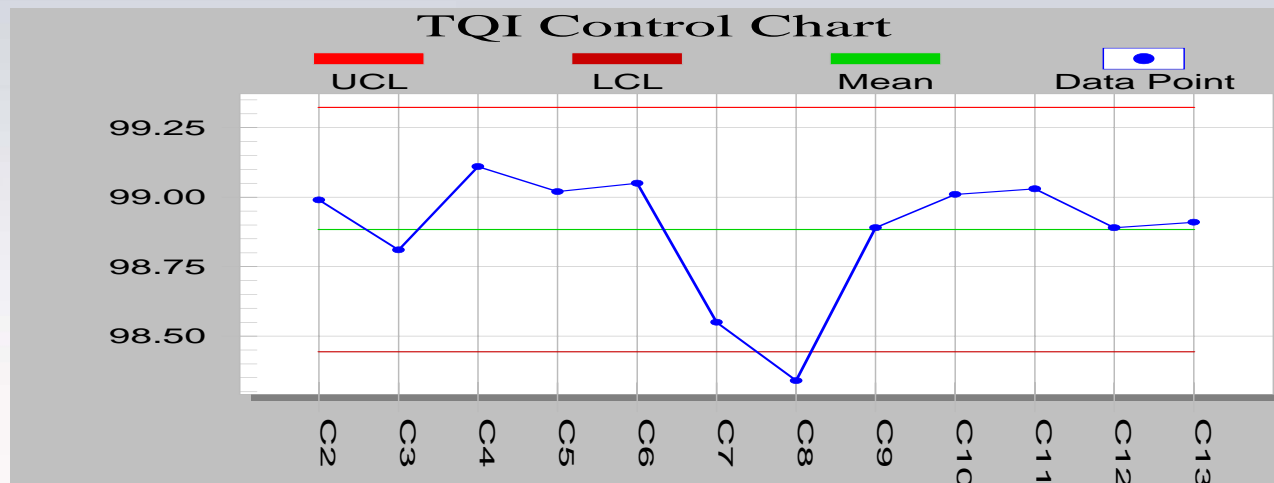
How to Set Up a Quality Measurement System – Stage 6, Process Improvement

- Take corrective actions (process improvement)

- Compare the TQI values before and after a process change to check for actual process improvement

Example of Process for Accepting or Rejecting a Translation Process

1) Determine and describe what your process actually is (NOT what you think it is or what the process should be)

2) Measure the quality you have now

3) Determine if you have special cases, and if so, eliminate them (what the special cases are can be seen through the use of control charts)

4) Once the process is in statistical control (i.e., any quality variance is not due to special cases)

5) Change the process to improve quality

6) Measure the new level of quality to determine the effectiveness of the changes to the process

Suggested process: calibration of a (generic) translation quality measurement tool

- *Have the sample translations (a suitable number of them) scored "by hand" by expert translators, so as to obtain a suitable range of evaluated samples, from very good to very bad.*
- **Importance of tightly defining the pool of reviewers**
- **Importance of instructions for reviewers**
- *Have <u>other</u> expert translators score the same tests, <u>but using the tool</u>*
- *On the basis of the results of the previous two steps, adjust the weights, types of errors, etc. in the tool until you are satisfied it is going to help in assessing translation quality - that is, until you are confident that trained evaluators are going to obtain with the tool consistent and reliable scores*
- **In doing this <u>remember</u> to remove from the kind of errors that <u>can</u> be assessed those that are controversial, i.e., those that lead to differences of opinion whether they are errors or not**
- *Finally adjust the tool so that it produces the range of error scores that is useful for your organization (e.g., if you want "0" or 100% as your perfect score)*
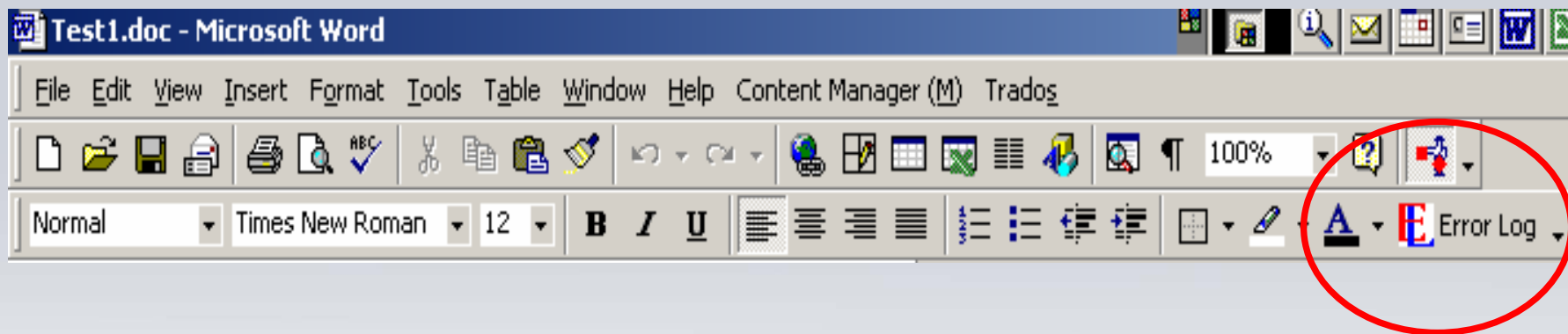
Calibration and Error Seeding

One of the things one can do to calibrate a translation quality measurement tool (or process) is to use <u>error seeding</u>: Not only to be able to estimate what percentage of errors is <u>not</u> discovered, but also in order to estimate how much variance there is in assessing the errors that do get discovered.
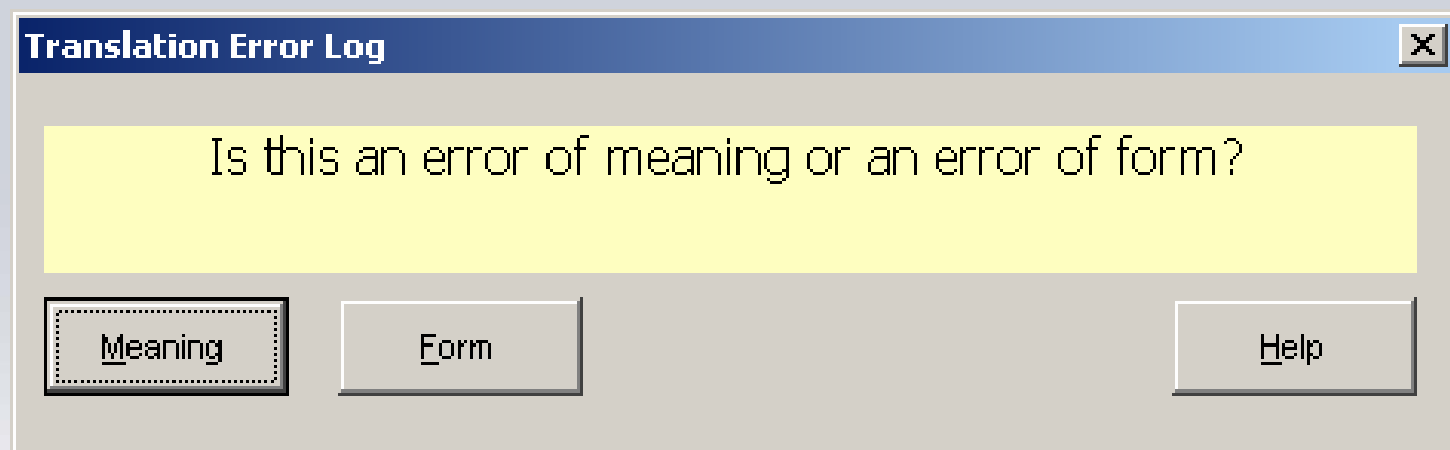
Translation Quality Measurement Tool

- The Translation Quality Measurement tool helps to measure process quality
- **It is NOT an editing tool, but it serves to measure whether a process is effective**
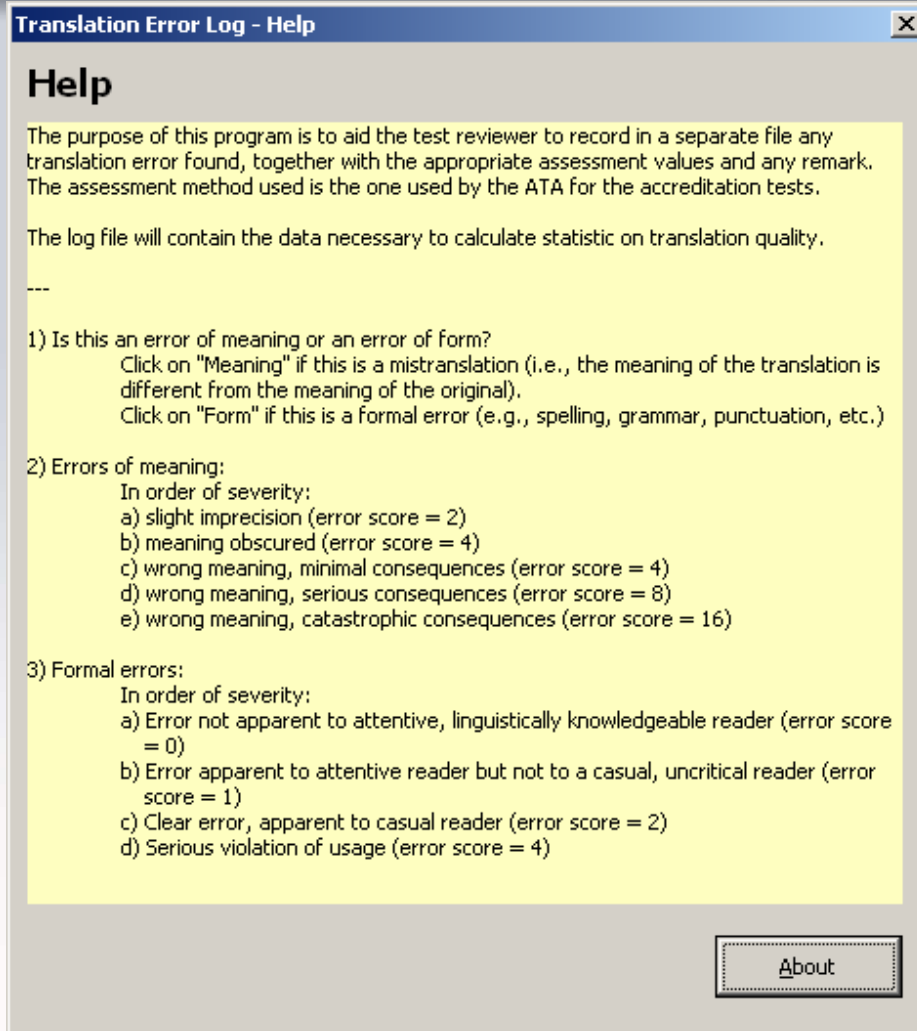
A TQI Tool Prototype

ATA Implementation

Translation Error Log

Is this an error of meaning or an error of form?

Meaning	Form		Help

ATA Implementation

Translation Error Log - Help

Help

The purpose of this program is to aid the test reviewer to record in a separate file any translation error found, together with the appropriate assessment values and any remark. The assessment method used is the one used by the ATA for the accreditation tests.

The log file will contain the data necessary to calculate statistic on translation quality.

---

1) Is this an error of meaning or an error of form?
   Click on "Meaning" if this is a mistranslation (i.e., the meaning of the translation is different from the meaning of the original).
   Click on "Form" if this is a formal error (e.g., spelling, grammar, punctuation, etc.)

2) Errors of meaning:
   In order of severity:
   a) slight imprecision (error score = 2)
   b) meaning obscured (error score = 4)
   c) wrong meaning, minimal consequences (error score = 4)
   d) wrong meaning, serious consequences (error score = 8)
   e) wrong meaning, catastrophic consequences (error score = 16)

3) Formal errors:
   In order of severity:
   a) Error not apparent to attentive, linguistically knowledgeable reader (error score = 0)
   b) Error apparent to attentive reader but not to a casual, uncritical reader (error score = 1)
   c) Clear error, apparent to casual reader (error score = 2)
   d) Serious violation of usage (error score = 4)

About

SAE Implementation (Modified)

Translation Error Log

Instructions

1 - Assign an error category to the error, and then the weight
2 - When an error is ambiguous, always choose the earliest error category (e.g., "grammatical error" before "misspelling")
3 - When in doubt, always choose "serious" over "minor"
4 - In case of errors due to legacy data (e.g. errors in the translation memory) or that were already present in the source language, please check the appropriate boxes

Error category
- (•) Wrong Term
- ( ) Incorrect meaning in TL
- ( ) Omission
- ( ) Grammatical error
- ( ) Misspelling
- ( ) Punctuation error
- ( ) Style error
- ( ) Miscellaneous error

Weight
- (•) Serious
- ( ) Minor

Legacy and SL errors
- [ ] Legacy Data error
- [ ] Source Language Text error

Remarks (optional):

[OK]   [Help]

SAE Implementation (Modified)

Translation Error Log - Help

Help

The purpose of this program is to aid the test reviewer to record in a separate file any translation error found, together with the appropriate assessment values and any remark. The assessment method used is based on the E-J2450 standard.

The log file will contain the data necessary to calculate statistic on translation quality.

About
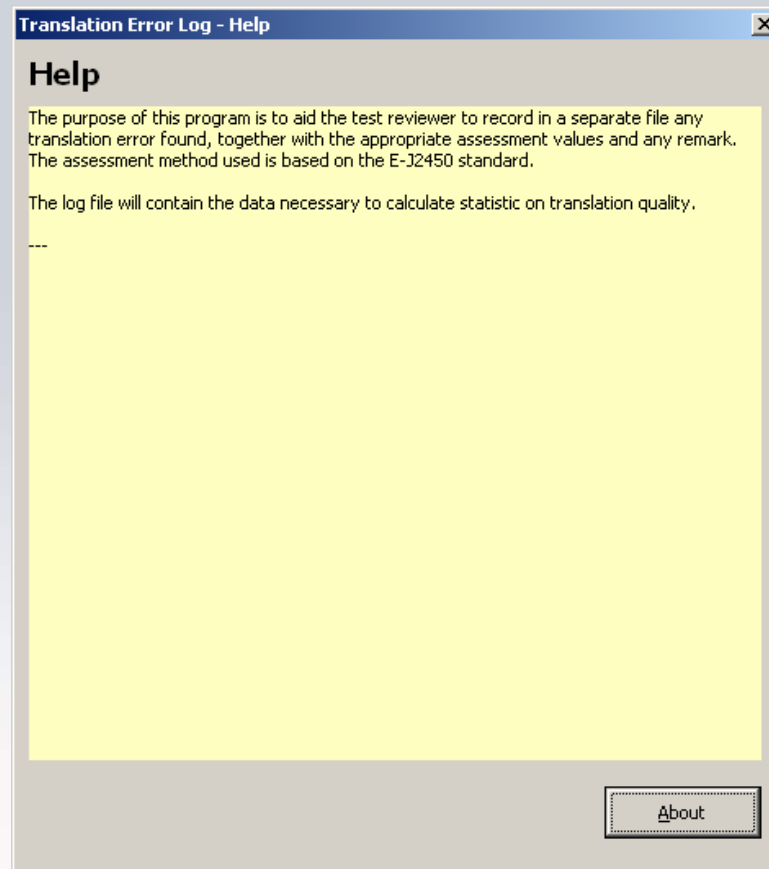
TQI Log

Error	Category	EP	Remarks	Bookmark	Path	File	Grader	Date
ls	Formal	2	irregular capitalization; should be is, not Is	x2_is	C:\Documents and Settings\RS1643403\Desktop\Quality Measurement	CoffeMakerTest.doc	RS1643403	11/1/2003
aluminium	Formal	1	British spelling; American English should be aluminum, not aluminium	x3_aluminium	C:\Documents and Settings\RS1643403\Desktop\Quality Measurement	CoffeMakerTest.doc	RS1643403	11/1/2003
food	Meaning	2	The container is not made to cook food, it is made to brew a beverage.	x5_food	C:\Documents and Settings\RS1643403\Desktop\Quality Measurement	CoffeMakerTest.doc	RS1643403	11/1/2003
right for the gas cooker, the electric plate and the pyroceram	Formal	1	A better phase might be: "acceptable for use on gas and electric stoves."	x6_right_for	C:\Documents and Settings\RS1643403\Desktop\Quality Measurement	CoffeMakerTest.doc	RS1643403	11/1/2003
pyroceram	Meaning	4	The word "pyroceram" is unknown to most English speakers.	x7_pyroceram	C:\Documents and Settings\RS1643403\Desktop\Quality Measurement	CoffeMakerTest.doc	RS1643403	11/1/2003
wash	Meaning	2	In English, the word "wash" typically means water and soap. The instructions specify only using water, so a better word choice would be "rinse."	x8_wash	C:\Documents and Settings\RS1643403\Desktop\Quality Measurement	CoffeMakerTest.doc	RS1643403	11/1/2003
trow	Meaning	2	The word "trow" is a misspelling of "throw."	x9_trow	C:\Documents and Settings\RS1643403\Desktop\Quality Measurement	CoffeMakerTest.doc	RS1643403	11/1/2003

total	14
N. of words	42
TQI	67%

Use of the Tool

- Use the tool to measure the effectiveness of quality control process

- Analyze the results obtained through the tool (control charts)

- If the process is NOT in statistical control
  - Discover special causes and deal with them appropriately
  - Remove them if they are negative
  - Incorporate them in process if they are positive

- Improve the process when it is in statistical control

How to Set Up a Quality Measurement System – Summary

1. Preparation
2. Calibration
3. Sampling
4. Measurement
5. Statistical Analysis
6. Process Improvement

Practical Recommendations

Importance of

- **Glossaries** (for terminology)

- **Style Guides** (for syntax)

- **Translation Instructions** (for special cases)

- **Protocols of Engagement** (regulating the treatment of errors/defects and defining the acceptance/rejection criteria)

- **Translation Guide for Customers** (including a detailed customer checklist to specify what is important and what is not)

Conclusions

- Desirability of **common standards** (see GAAP - Generally Accepted Accounting Principles)
  - It is not possible to directly compare different quality initiatives
  - A common standard would still permit assigning different weights to different categories but in a much more transparent and comparable way
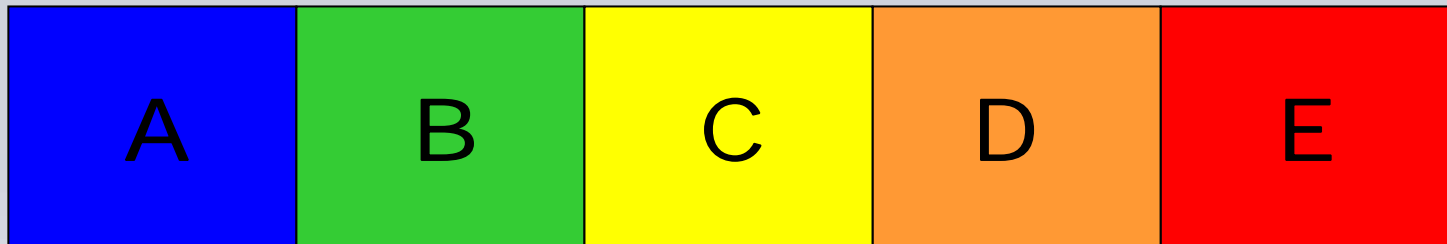
Translation Quality Scale

Quality Continuum

Contacts

Riccardo_Schiaffino@aliquantum.biz – Aliquantum, Inc.

Franco.Zearo@lionbridge.com – Lionbridge Technologies, Inc.

For more information about translation issues, visit

TranslationQuality.com

Biographical Notes on the Authors

Riccardo Schiaffino

Riccardo Schiaffino worked as translator, translation manager and special software translation project lead for a major software company, and now leads a small company he established with a few experienced colleagues. As a translation manager, Riccardo worked on the improvement of translation quality and on translation quality metrics and tools. He holds an MA degree in Translation, and has been working in translations for over 20 years, first in Italy and then in the U.S. Riccardo is ATA accredited.

Franco Pietro Zearo

Franco Pietro Zearo is a project manager with Lionbridge Technologies in Boulder, Colorado. He holds a degree in translation from the Advanced School of Modern Languages for Translators and Interpreters at the University of Trieste, Italy, and earned an MBA from the University of Phoenix. Before joining Lionbridge in 1996, he worked as a freelance technical translator in Italian, English, and Russian. At Lionbridge, he has held positions in translation, localization analysis, presales, and cultural and globalization consulting. He has been responsible for translation quality on numerous projects for many Fortune 500 clients. In his previous role as senior technical translator, he helped define best practices for the translation department.