Linguistic Quality Assurance (LQA) as One of the Key Parts of the Localization Process

An Inside Look
What is LQA

• LQA refers to assessing Linguistic Quality of [translated] materials based on:
  – International & industry-wide standards
  – Client’s standards, requirements & guidelines, including
    • Approved Terminology & Style Guides
    • Agreed Metrics & Quality Criteria

• LQA is primarily expected to check *how* good or bad final materials are
What LQA Is NOT

• “Hybrid” approaches do not work well:
  – LQA & Editing
    • Feedback not getting through, improvements unlikely
  – LQA & Functional Testing
    • Linguists are typically not too good at spotting or analyzing technical bugs, let alone SW configurations
    • Dedicated testers do this better and faster across all languages
  – Translation & LQA in one box (at one vendor)
    • High probability of artificial adjustment or incomplete logging of LQA results

(C) 2010, Logrus International
Why Is LQA Necessary

• Peace of mind
  – Independent, third-party examination of materials
  – Doesn’t take too long, is not too expensive
  – Avoiding costly errors at a fraction of the price

• Means of Vendor Selection and Translation Process Quality Control
  – Consistent failures might mean something is wrong with the vendor or the process

(C) 2010, Logrus International
Why Is LQA Necessary (II)

• Offsetting/preventing negative effect on quality produced by latest trends in translation

*Getting a “Well-Disguised Less” for Less*

– Declining rates
– Wider MT Application without due process
– Vendor consolidation
– Work fragmentation
– Extensive mark-up language usage
– Unlimited recycling
How Can We Measure Quality?

• Objective Criteria
  – Recognized and Univocal
  – Easily Applicable (no grey areas)
    • Violations/Deviations can be clearly described
    • Proof is universal. Understanding its essence does not require knowledge of the language
  – Typical Examples:
    • Language (Spelling & Grammar)
    • Correct References, No (Over-/Under-)Translations
    • Country & Other Standards
    • Terminology
    • Style Guide & Explicit Client’s Guidelines
How Can We Measure Quality? (II)

• Expert Opinion-Based (Semi-Objective) Intelligibility, Adequacy, Equivalence, Fluency…
  – An Expert Panel would produce a normal Opinion Curve around the Average Value

• Subjective Criteria
  – Preferential, taste-based, obscure arguments: “I don’t like it”, “This is bad”, “Poor style” “That way it sounds better”, …
  – One can’t explain what’s wrong and why
  – The feedback is not well structured
  – An Expert Panel would produce a “White Noise”-type Spectrum with no pattern

(C) 2010, Logrus International
Is LQA as Un-Objective as It Seems?

- Apply Expert Opinion-Based (Semi-Objective) Criteria as PASS/FAIL Ones
  - Define Acceptance Threshold
  - Select Criteria & Define Grading System
    These are NOT as accurate and can’t be naturally combined with Objective ones in a formula

- Select Objective Criteria
  - Assign Weights & Define Show-Stoppers

- Ignore Subjective Complaints

- Use Representative Sampling – See Further!
Apply Expert Opinion-Based Criteria
  • Generate an integral Expert-Opinion Based Rating
  • What Acceptance Threshold really means?
  • FAIL everything that falls below this threshold

Apply Objective Criteria to Whatever Is Left
  – Integral Expert Rating (e.g., 7.8 out of 10)
  – Integral Objective Rating (e.g., 8.6 out of 10)

Can be combined into a single rating if
  – Weight assigned to the Expert part is below 30%

Works perfectly for MT (adjusting thresholds)
LQA Results and the Turtle
How Much Should Be Reviewed?

- Do we need to check everything?
- If a certain % is sufficient,
  - Is there a scientific approach to selecting the overall volume to be QA-d?
  - What is the “magic” sampling scope that would guarantee peace of mind?
- How can we produce reliable and representative QA results?
LQA & Opinion Polls

• We can only poll a limited number of people  
  OR Review a limited number of words

• After polling N people (58% said "YES")  
  OR Reviewing N words (“PASS” for 85% of all segments)  
  we need to assess the CREDIBILITY of the result  
  (Or find N that guarantees reasonable credibility)

• We have to use the so-called  
  Confidence Level (CL) & Confidence Interval (CI):  
  “… We can assert with 95% confidence (Confidence Level) that 58% of the population will vote for X.  
  The margin of error (Confidence Interval) is 5% for this survey…”
LQA Specifics: Setting CL & CI

• LOW Error Levels expected (Typically less than 1 error is allowed per 100 words)
• HIGH Precision Required. Opinion Poll Analogy:
  - Will a marginal “United Incompetence Party” get parliament representation given the 1% Election Threshold?
• Margin of error (CI) must fall WELL BELOW the allowed/expected Translation Error Levels, i.e. below 1%
• Required Sample Size (QA-d Volume) depends on:
  • Overall Volume (Population)
  • Confidence Interval (Margin of Error): << 1%
  • Confidence Level (Reliability): Typically set at 90-95%
How Fastidious Are You?

% to Be Checked

Sample Size (Words)

Total Volume, Words

95% Confidence Level

% of Total - 0.25% CI

Sample Size - 0.25% CI

% of Total - 0.5% CI

Sample Size - 0.5% CI

% of Total - 0.125% CI

Sample Size - 0.125% CI
LQA Sampling Summary

- Not applicable to crowdsourcing-type projects
- Optimal Sampling: Random Selection, No Exclusions! Recommended: One-page (250 words) pieces
  - Size convenient for reviewers
  - Big enough to make conclusions about adequacy, fluency, etc.
  - Small enough to provide representative stats (10-30 pages/person)
- ALWAYS Check ALL Priority/Exposed Pieces in Full
- MUCH LESS effort required to check for a LEMON: CI = 1%, Sample Size = 10 K words

<table>
<thead>
<tr>
<th>Volume</th>
<th>Sample Size / Check</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 10 K words</td>
<td>100%</td>
</tr>
<tr>
<td>20 -&gt; 200 K words</td>
<td>85% -&gt; 45% (+/- 25%)</td>
</tr>
<tr>
<td>&gt; 300 K words</td>
<td>100-150 K words (UP TO 3 TIMES More/Less)</td>
</tr>
</tbody>
</table>
LQA Specifics: Process

• Mostly Small Projects with Short Turnaround Times
  – Productivity: Higher than Translation; Volumes: Smaller
  – HUNDREDS of operations per each small project
    • Hand-off
    • Confirmation [Finding Alternative Reviewer(s)]
    • Sending materials
    • Answering questions & Provide clarifications
    • Getting reviews back
    • Checking review technical quality, completeness, consistency, etc.
    • Hand-back
  – Schedule typically squeezed in case of translation slips
    • Ramp-up and Turnaround times limited by time zone spread and holidays
  – An Automated Workflow Portal is a Must
    • Job Costing & Timing Estimates
    • Order Placement
    • 24 x 7 Status
    • Automated Notifications

(C) 2010, Logrus International
LQA Specifics: Review Level

- QA is Neither Translation Nor Editing!
  - Special Training Required
  - No fixes, no “improvements”…
  - Formalized, LQA-specific requirements and expectations
  - Mandatory use of client- & project-specific guidelines & reference materials for *each* job, however different/weird these might seem
  - Formalized feedback forms
  - Strict evaluation metrics
  - Suppression of emotions
  - Ignoring OR Imposing style-related considerations
  - Reconciliation discussions

(C) 2010, Logrus International
LQA Cost & Pricing Model

• Small Projects = High Managerial Overhead
• Approach based on Hourly Linguistic rates is deficient
  – Managerial costs exceed linguistic costs on small projects
  – Either losing money on small jobs or overcharging clients on big ones
• A more cost-oriented and transparent approach is needed
  – Separating PM and Language Work. (Both are charged by hour)
  – PM: Volume based on No. of Languages & Word counts
    • PM hourly cost is language-independent
  – Linguistic work: Volume based on Word count & Productivity
  – Utilizing a costing formula combining PM & Language work
    • Starts with non-zero, small PM cost for tiny projects (Minimal Fee equivalent)
    • PM hours proportional to number of languages
    • Grows linearly but slowly with volume
    • At high volume language costs dominate

(C) 2010, Logrus International
LQA: Major Challenges

• Productivity expectations and “The Big Disappointment”
  – Average productivities based on GOOD translation quality
  – In the case of poor quality it drops, and more time is needed
  – What’s to be done if allocated time has been spent
    • Stoppage flags in case of poor quality
    • Stopping QA when the agreed hourly limit is reached
LQA: Major Challenges (II)

- Potential conflicts and never-ending disputes between parties
  - Vendors tend to dispute EVERYTHING, even if there’s no chance to change final QA results
    - Lengthy, time-consuming discussions without any results
    - Huge and costly increase in PM time/overhead
  - Eliminating inconsistencies in file versions, reference materials and guidelines
  - Limiting the number of LQA dispute iterations is a must
    - There should be a formal way to break the vicious circle
  - Vendors should not dispute results unless
    - LQA feedback contains serious mistakes
    - Ratings might be seriously changed as a result

(C) 2010, Logrus International
LQA: Major Challenges (III)

• Objectivity and adequacy of reviews
  – Reviewers need to undergo general LQA training & custom training for each client
  – LQA review quality itself has to be checked on a regular basis

• Choice of balanced evaluation criteria
  – Avoiding sharp dependencies on any particular factors
  – Limiting subjectivity: PASS/FAIL approach to Expert Opinions

• Mismatching expectations of parties involved

• Scheduling & timing problems
  – LQA follows translation, almost always on a critical path
Is It Doable and Useful?

YES!