STANDARDIZING QUALITY ASSESSMENT FOR THE MULTILINGUAL WEB

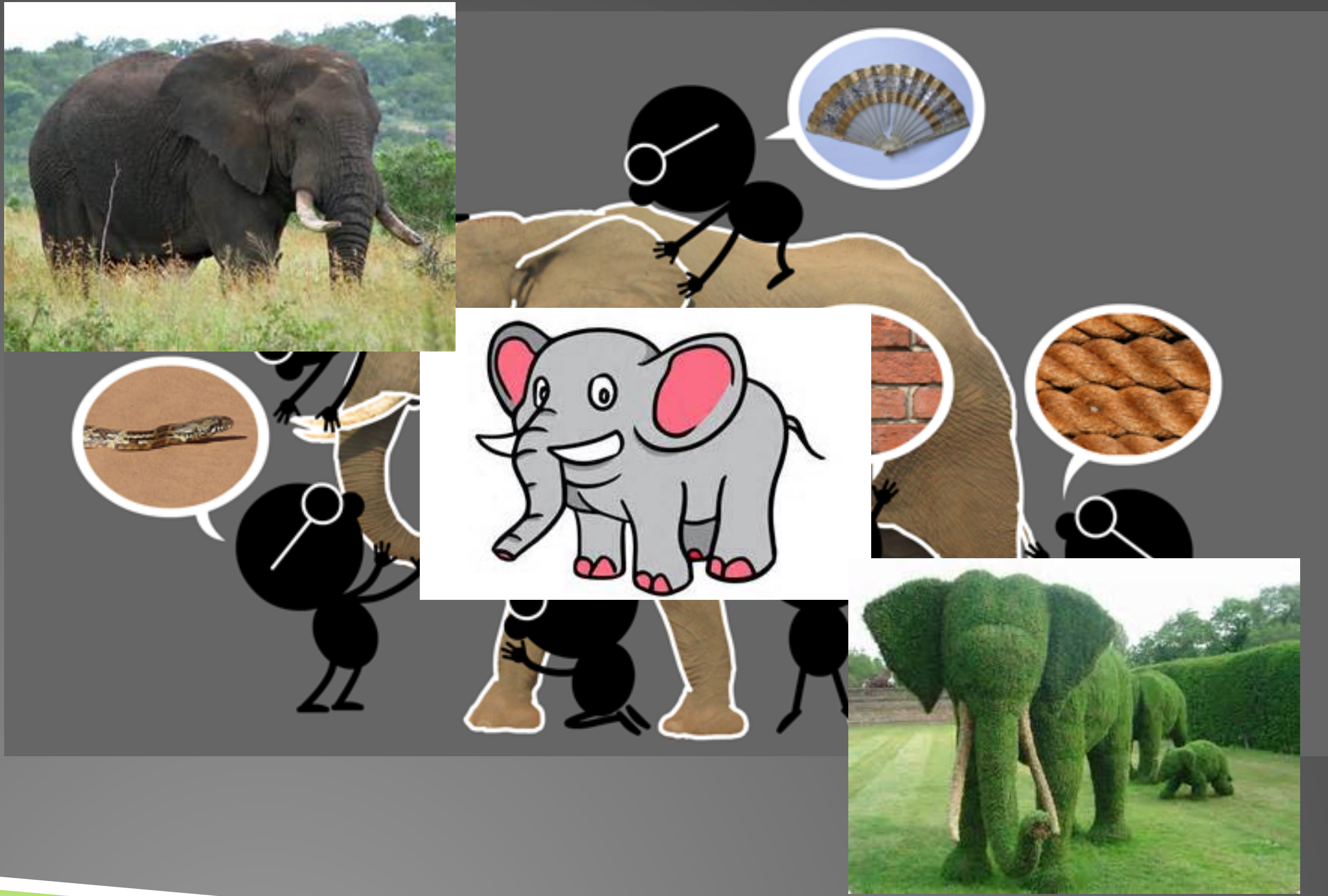
Leonid Glazychev, Ph.D., CEO
Logrus International Corporation

ASTM STANDARD PROPOSAL WK46397

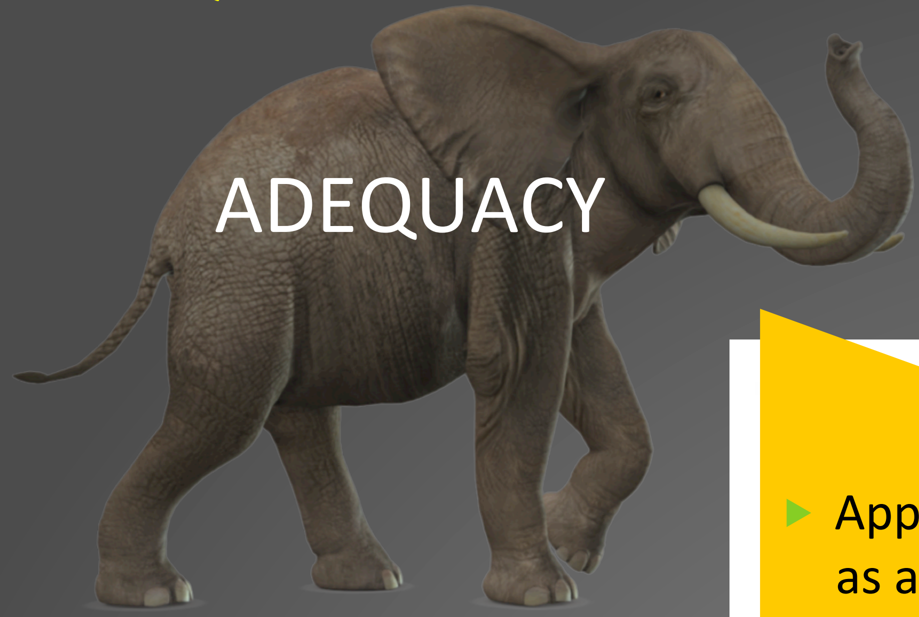
- RATIONALE
  - Standards crucial for all stages of content production
    - Including quality assessment of multilingual materials
  - No methodology or metrics for public Language Quality Assurance (LQA)
  - Executive Order 13166: http://www.lep.gov/, http://www.justice.gov/crt/about/cor/13166.php "Improving Access to Services for Persons with Limited English Proficiency“, reaffirmed in 2011
- WK46397
  - “Development of a complete methodology, including a simplified quality metric, for crowd-sourced expert language quality assessment targeted at nonprofit web sites and other documents of public interest.”
- PRIMARY GOAL: A simplified quality assessment standard
  - Quick, inexpensive and reliable initial assessment
  - Reviewing big, highly visible resources designated for wide public use
    - Large and significantly diverse target audience
    - Limited review capabilities and/or budget

EMPHASIS ON HOLISTIC ASSESSMENT

- The whole is always more important to us than its constituents
- Object properties can’t be fully revealed or described based on its parts alone
- Quality assurance cannot be complete or accurate if there is no way of making holistic evaluations



THE QUALITY TRIANGLE

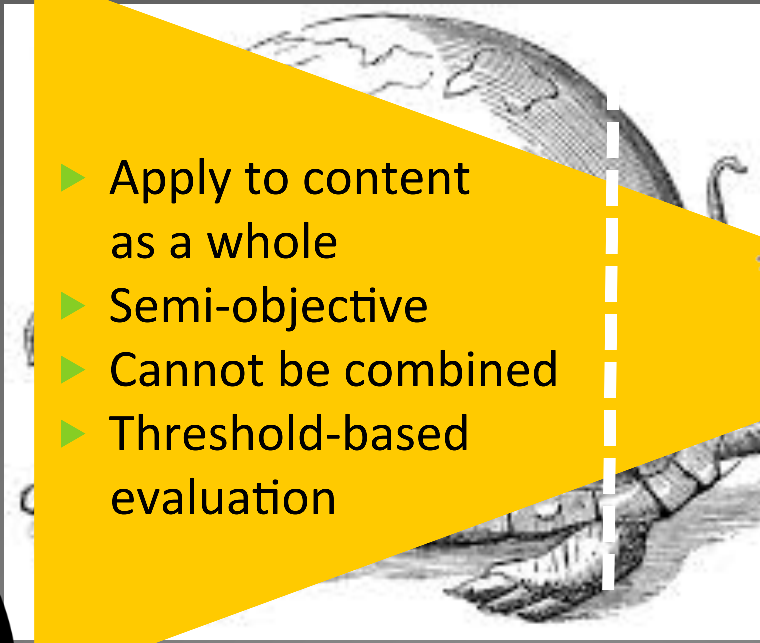


ADEQUACY

Acceptance Threshold

Universal, 3D quality picture

- Same approach
- Any issue catalogue
- Only expectations vary

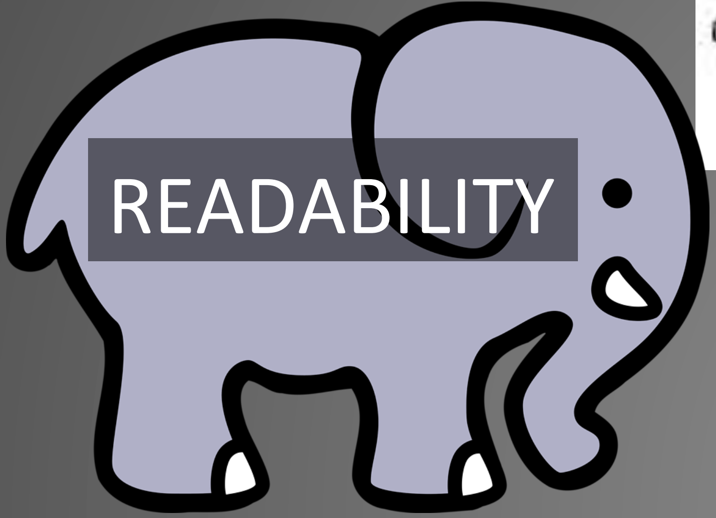


- Apply to content as a whole
- Semi-objective
- Cannot be combined
- Threshold-based evaluation

ATOMISTIC QUALITY

$Q_A = \sum_{i=1}^{\infty} (W_i N_i)/V$

HOLISTIC FACTORS

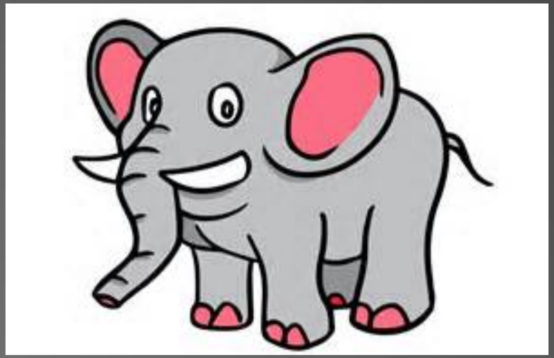


READABILITY

Acceptance Threshold

ATOMISTIC QUALITY

- Measured in ALL existing quality metrics
- Opposite to holistic
- Applies to
  - Quality issues at the “atomic” level of the content (vs. holistic)
  - Sentences, strings, translation units, …
- Includes issues like
  - Terminology inconsistency or deviations
  - Style guides, country standards
  - Tags, placeholders
  - Formatting
  - …
- Complements holistic usability/quality evaluation
- Example of a comprehensive issue framework
  - MQM: http://www.qt21.eu/launchpad/content/multidimensional-quality-metrics



THE PRICE OF OBJECTIVITY

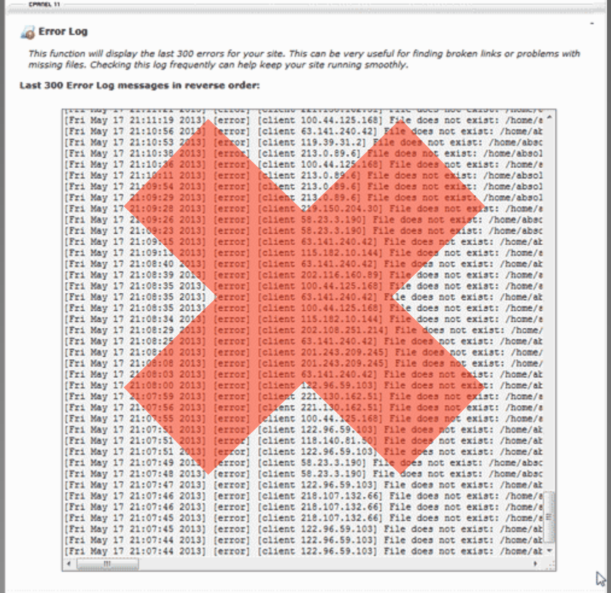
- Objective = Universal issue nature
  - Explanation doesn't require language knowledge
- No dependence on the reviewer's personality
  - A typo is still a typo
  - An error in country standards is still an error anyway
  - Everything depends on issue classification and the weighting system
- Price of objectivity
  - Comprehensive and clear issue classification
  - Availability of all ancillary materials
    - Glossaries, style guides, special requirements, etc.
  - Professional reviewers
  - Extensive language quality assurance (LQA) training
  - Detailed issue logging
  - Issue reconciliation with translators
  - Time and cost
  - **Otherwise the assessment is subjective and inaccurate!**

THE CHALLENGE

THE LIMITATIONS

- Cannot expect serious preparation
  - Minimal/no reviewer training
  - Just explain the task in the simplest terms possible
- No thorough issue catalogues/quality frameworks
  - Unless completely trivial
- No serious quality issue logging
  - Just ask to provide typical examples
  - Make the feedback form simple and short

- **OUT OF THE QUESTION:**
  - **Complicated requirements**
  - **Strict definitions**
  - **Quality frameworks**
  - **Special rules, etc.**



Error Log

This function will display the last 300 errors for your site. This can be very useful for finding broken links or problems with missing files. Checking this log frequently can help keep your site running smoothly.

Last 300 Error Log messages in reverse order:

```
[Fri May 17 21:11:19 2013] [error] [client 100.44.125.168] File does not exist: /home/a
[Fri May 17 21:10:56 2013] [error] [client 63.141.240.42] File does not exist: /home/ab
[Fri May 17 21:10:53 2013] [error] [client 119.39.31.2] File does not exist: /home/abs
[Fri May 17 21:10:38 2013] [error] [client 213.0.89.6] File does not exist: /home/absol
[Fri May 17 21:10:36 2013] [error] [client 100.44.125.168] File does not exist: /home/a
[Fri May 17 21:10:11 2013] [error] [client 213.0.89.6] File does not exist: /home/absol
[Fri May 17 21:09:54 2013] [error] [client 213.0.89.6] File does not exist: /home/absol
[Fri May 17 21:09:29 2013] [error] [client 213.0.89.6] File does not exist: /home/absol
[Fri May 17 21:09:28 2013] [error] [client 219.150.204.30] File does not exist: /home/a
[Fri May 17 21:09:26 2013] [error] [client 58.23.3.190] File does not exist: /home/abs
[Fri May 17 21:09:23 2013] [error] [client 58.23.3.190] File does not exist: /home/abs
[Fri May 17 21:09:15 2013] [error] [client 63.141.240.42] File does not exist: /home/ab
[Fri May 17 21:09:13 2013] [error] [client 115.182.10.144] File does not exist: /home/a
[Fri May 17 21:08:40 2013] [error] [client 63.141.240.42] File does not exist: /home/ab
[Fri May 17 21:08:39 2013] [error] [client 202.116.160.89] File does not exist: /home/a
[Fri May 17 21:08:35 2013] [error] [client 100.44.125.168] File does not exist: /home/a
[Fri May 17 21:08:35 2013] [error] [client 63.141.240.42] File does not exist: /home/ab
[Fri May 17 21:08:35 2013] [error] [client 100.44.125.168] File does not exist: /home/a
[Fri May 17 21:08:34 2013] [error] [client 115.182.10.144] File does not exist: /home/a
[Fri May 17 21:08:29 2013] [error] [client 202.108.251.214] File does not exist: /home/
[Fri May 17 21:08:25 2013] [error] [client 63.141.240.42] File does not exist: /home/ab
[Fri May 17 21:08:10 2013] [error] [client 201.243.209.245] File does not exist: /home/
[Fri May 17 21:08:08 2013] [error] [client 201.243.209.245] File does not exist: /home/
[Fri May 17 21:08:03 2013] [error] [client 63.141.240.42] File does not exist: /home/ab
[Fri May 17 21:08:00 2013] [error] [client 122.96.59.103] File does not exist: /home/ab
[Fri May 17 21:07:59 2013] [error] [client 221.130.162.51] File does not exist: /home/a
[Fri May 17 21:07:56 2013] [error] [client 221.130.162.51] File does not exist: /home/a
[Fri May 17 21:07:55 2013] [error] [client 100.44.125.168] File does not exist: /home/a
[Fri May 17 21:07:51 2013] [error] [client 122.96.59.103] File does not exist: /home/ab
[Fri May 17 21:07:51 2013] [error] [client 118.140.81.50] File does not exist: /home/ab
[Fri May 17 21:07:51 2013] [error] [client 122.96.59.103] File does not exist: /home/ab
[Fri May 17 21:07:49 2013] [error] [client 58.23.3.190] File does not exist: /home/abs
[Fri May 17 21:07:48 2013] [error] [client 58.23.3.190] File does not exist: /home/abs
[Fri May 17 21:07:47 2013] [error] [client 122.96.59.103] File does not exist: /home/ab
[Fri May 17 21:07:46 2013] [error] [client 218.107.132.66] File does not exist: /home/a
[Fri May 17 21:07:46 2013] [error] [client 218.107.132.66] File does not exist: /home/a
[Fri May 17 21:07:45 2013] [error] [client 218.107.132.66] File does not exist: /home/a
[Fri May 17 21:07:45 2013] [error] [client 122.96.59.103] File does not exist: /home/ab
[Fri May 17 21:07:44 2013] [error] [client 122.96.59.103] File does not exist: /home/ab
[Fri May 17 21:07:44 2013] [error] [client 122.96.59.103] File does not exist: /home/ab
```

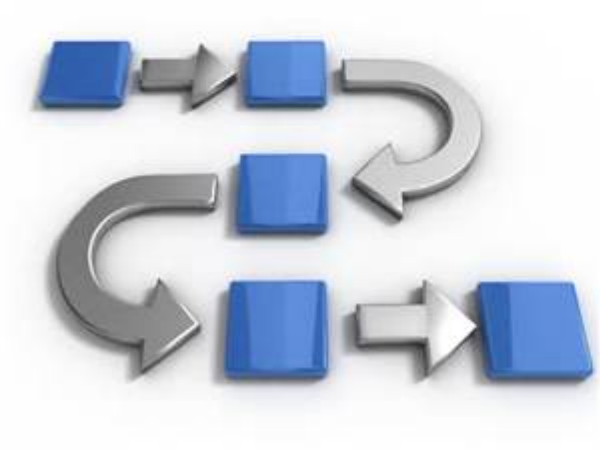
DEVELOPING A CROWDSOURCING-BASED QUALITY STANDARD

- Defining all three cornerstones

General Approach/Methodology

Quality Metric

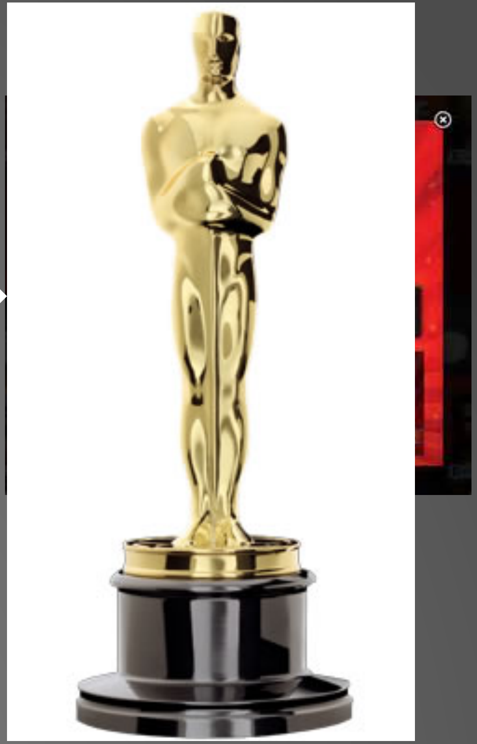
Process



(C) 2015
Logrus International

GENERAL APPROACH AND METHODOLOGY

- Simplified methodology
  - Focusing on holistic evaluations
- Objectivity and accuracy gained through statistics
  - Meaningful averages and standard deviations
  - Multiple people reviewing the same piece
  - Essential to collect sufficient statistics
- Limit contributors to language professionals only

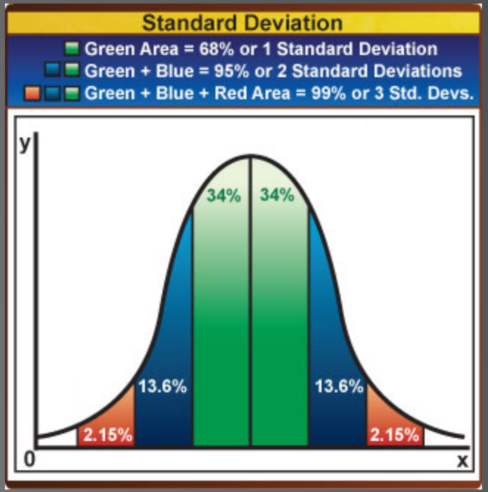


SIMPLIFIED QUALITY SQUARE METRIC

- The Quality Square approach
- Simplified form, no detailed issue definitions or formal requirements
- Four ratings for each text on a 0-10 scale
  - The number of major (showstopper) errors
    - 0 => 10
    - 1 => 5
    - 2 or more => 0
  - Holistic translation readability
    - 0 = Completely unreadable/incomprehensible
    - 10 = Perfectly intelligible and readable text
  - Holistic translation adequacy
    - 0 = Completely inadequate
    - 10 = Perfectly conveyed meaning
  - Atomistic quality
    - 0 = Overabundance of atomistic-level errors
    - 10 = Completely error-free text
  - <u>**A brief explanation required in each case**</u>

QUALITY ASSURANCE PROCESS

- Clear and brief LQA review scope
- Translated content frozen
- Online portal with project description/scope definition
- Pre-process results
- Individual list of pre-processing checks for each project
- Calculating median ratings and standard deviations
- Comparing all ratings against pre-defined thresholds



EXPECTATIONS

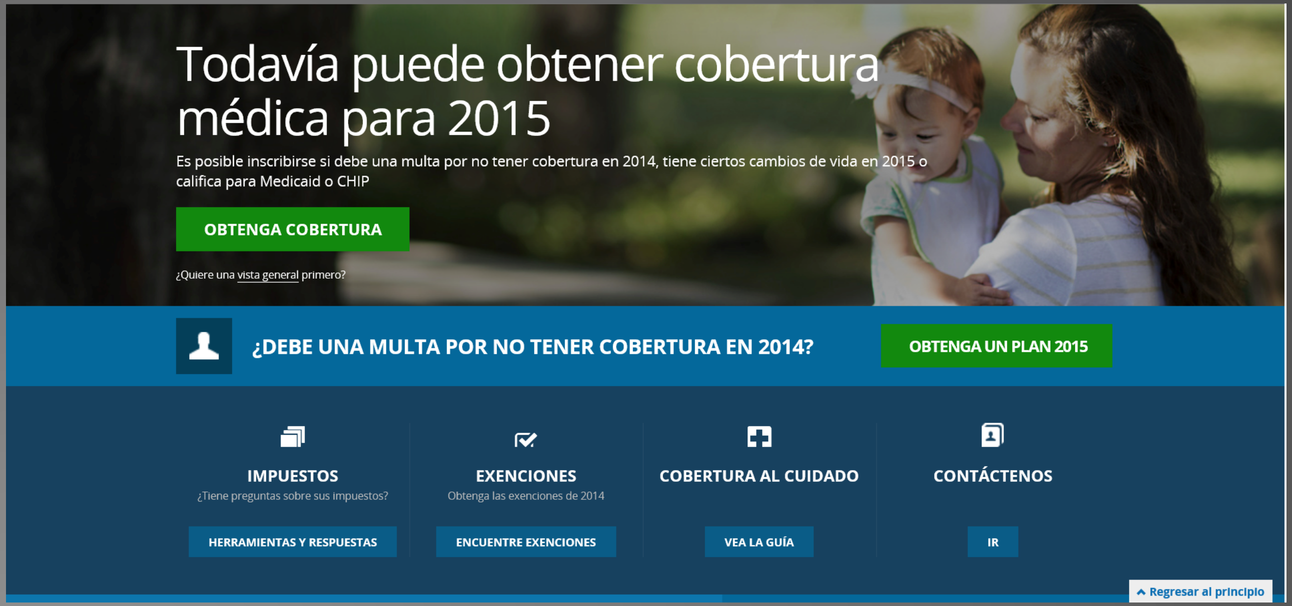
- Reliable, statistically sound high-level LQA results in the crowdsourcing environment
- Cannot serve as a valid replacement for professional LQAs

Untrained translators or linguists	Specially trained professionals
Almost no formal criteria	Extensive and well-defined formal criteria
General criteria	Criteria fine-tuned to the client's requirements
Minimal accuracy and consistency	High level of sophistication, accuracy, and consistency

- Obtaining quick results at a minimal (or zero) cost
  - Getting a rough evaluation of translation quality
  - Reveal significant problems
  - Assess the need for a professional LQA
- Acceptance thresholds replaced by "alarm-raising" ones

CASE-STUDY: US ACA SPANISH WEBSITE REVIEW

- Originally requested directly by the US government
  - Affordable Care Act Spanish-Language Website: www.CuidadoDeSalud.gov
- Carried out free of charge by Logrus International for GALA
  - Globalization and Localization Association, www.gala-global.org
- Logrus developed and provided methodology
- Logrus organized the review and provided analytics
- Volunteer effort, crowdsourcing-based approach



Todavía puede obtener cobertura médica para 2015

Es posible inscribirse si debe una multa por no tener cobertura en 2014, tiene ciertos cambios de vida en 2015 o califica para Medicaid o CHIP

¿Quiere una vista general primero?

¿DEBE UNA MULTA POR NO TENER COBERTURA EN 2014?

IMPUESTOS ¿Tiene preguntas sobre sus impuestos?

EXENCIONES Obtenga las exenciones de 2014

PROCESS ORGANIZATION

- Strictly following the process described earlier
- Simplified Quality Square methodology applied
  - Major errors (10 = None, 0 = More than 2)
  - Readability (0 - 10)
  - Adequacy (0 - 10)
  - Atomistic (0 – 10)
- 18 contributors chosen among language professionals only
- Mini-portal for participants
  - Self-registration
  - Brief error category definitions
  - Entering ratings and comments
- Comprehensive data pre-processing, discarding:
  - Standalone “perfect” (10 out of 10) evaluations
  - Marginally high or low ratings with no explanations
  - Skewed ratings caused by reviewer errors

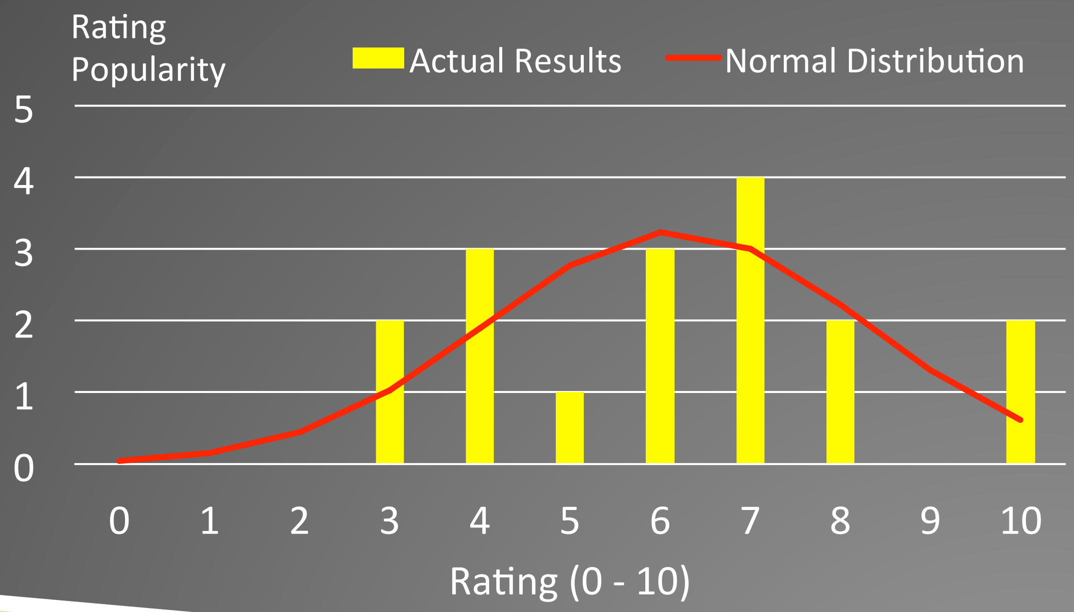
PROJECT SPECIFICS

- Target language specifics
  - Most translation and LQA tasks target a specific region
    - Latin America (LatAm), Argentina, Mexico, Spain…
  - Each reviewer had a particular language “flavor” in mind
  - Target audience = Spanish-speaking population in the US
    - People with various backgrounds
    - Speaking a wide variety of Spanish, or even “Spanglish”
  - Most neutral and universal translation not sounding natural to some native speakers
- Understanding the review scope
  - Some “major errors” were functional issues beyond the LQA scope
    - Navigating health insurance plans and prices in English
    - Spelling errors in responses obtained through the chat feature
- Disregarded during pre-processing
  - Targeting translation quality alone, not portal usability or functionality

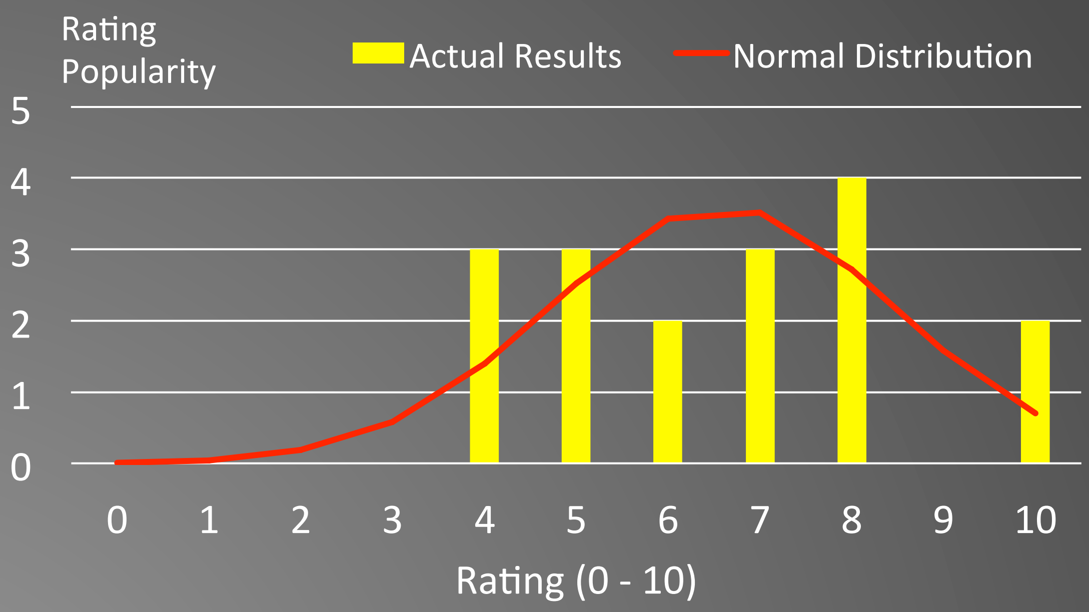
MAJOR ERRORS – READABILITY – ADEQUACY

- Major errors: None (11), More than 2 (7), 1 grade ignored
- Readability and Adequacy
  - **YOUR** reviewer could contribute to ANY of the bars
  - Only threshold-based criteria really work

Readability. Mean value: 6.2, Std. Deviation: 2.1



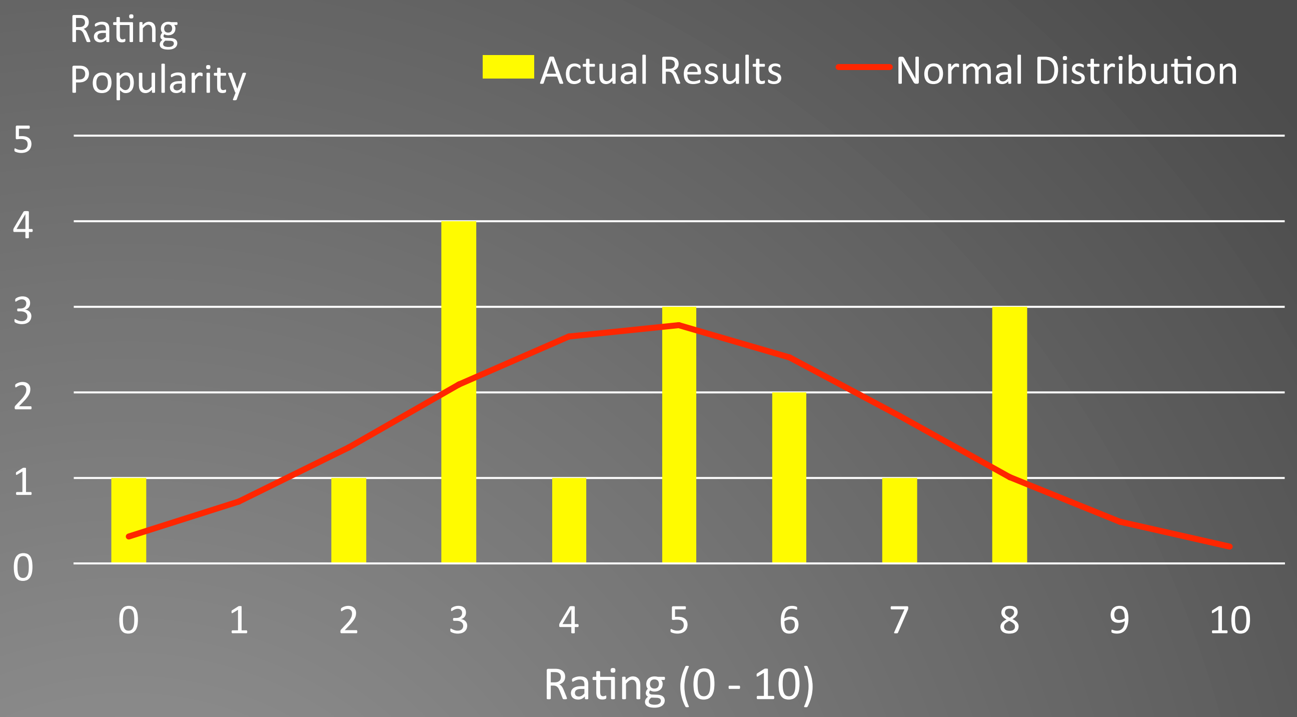
Adequacy. Mean value: 6.6, Std. Deviation: 1.9



ATOMISTIC QUALITY

- Biggest opinion spread
  - Illustrates the gap between professional and crowd-sourced work
  - No detailed criteria or training
  - Should be the most objective factor ☺
- “Mechanical” stats
  - Mean value: 5.4
  - Standard deviation: 2.8
- Adjusted stats
  - Mean value: 4.7
  - Standard deviation: 2.4

Atomistic Quality. Mean value: 4.7, Std. Deviation: 2.4



(C) 2015
Logrus International

METRIC AND PROCESS SUMMARY

- Both holistic Readability and Adequacy results can be relied upon
- Good basis for assessing overall translation quality
- Judgment about the presence of Showstopper errors is convincing
- Atomistic quality assessment is not accurate enough
  - Gives a good general idea of the pervasiveness of non-critical, atomistic-level errors

- Major crowdsourcing LQA results look trustworthy and consistent
- A reliable high-level picture of translation quality
- Experimental proof that the whole model works
  - Even in the relatively extreme crowdsourcing environment