Bad to the Bone: AI-Enabled SmartLQA
Alex Yanishevsky

Director, AI Deployments
Welocalize
SmartLQA Agenda

WHAT IS IT?

WHEN IS IT USED?

HOW IS IT USED?

WHAT’S NEXT?
What is it?

Methodology to inform strategic global content business decisions through AI

- SOURCE SUITABILITY
- PREDICT AT-RISK CONTENT
- “SPENDING SMART” VIA TARGETED LQA
- MTQE CORRELATION
- PE DISTANCE CORRELATION
What is it?

AI-Driven Quality Management
Inform data-driven content decisions through AI

SOURCE SUITABILITY
AI can identify errors in poor source content and predict ‘at-risk’ content:
- Content written by non-native authors
- Content created by technical specialists for a non-technical audience
- Dated content not adhering to brand tone and voice

Does the source content need to be re-written before translation?
What is it?

AI-Driven Quality Management

Inform data-driven content decisions through AI

TARGET SUITABILITY

• Does the translation deviate from previous style?
• Does the translation introduce unnecessary complexity?

Does the target need go through LQA for data-driven checks and corrections?
AI-Driven Quality Management

Inform data-driven content decisions through AI

AI-DRIVEN LQA + MT RETRAINING

- Targeted “SmartLQA” focuses on problematic files and segments within them
- Data can be used to **retrain engines (dynamically)**
When Is It Used?
Where this fits into the Content Lifecycle

Source
Source Suitability
Target
Target Suitability
AI-driven LQA
MT retraining capabilities (dynamic)
How is it Used?

Configuring Thresholds

1. Based on average plus standard deviation(s)
2. Relative measure
3. Captures outliers for that specific domain/product
How is it Used?

Configuring Thresholds

- Based on average plus standard deviation(s)
- Relative measure
- Captures outliers for that specific domain/product
How is it Used?

Identifying Salient Features

1. Parts of speech such as adjectives, nouns, proper nouns, numbers
2. Adjective/noun density
3. Long words, complex words, short and long sentences
4. Stylistic similarity/dissimilarity
5. Readability and complexity metrics
6. Correlations to PE Distance and MT Quality Estimation metrics
How is it Used?
Identifying Salient Features

- Parts of speech such as adjectives, nouns, proper nouns, numbers
- Adjective/noun density
- Long words, complex words, short and long sentences
- Stylistic similarity/dissimilarity
- Readability and complexity metrics
- Correlations to PE Distance and MT Quality Estimation metrics
How is it Used? **Source Suitability**

### POSSIBLE REMEDIES

- Don't run the project till source is improved
- Route to transcreation, human translation, different MT engines
- Alert of higher LQA risk to all production people (PM, linguists, LQA)
How is it Used? **Source Query Analysis**

**PROCESS**
- Analyzed over historical 600 segments for potential DNT
- Analyzed almost historical 400 segments for source ambiguity and meaning (almost 200 for each category)
- Identified thresholds for each category
- Ran thresholds for all categories and identified over 400 potential queries
- Savings of 6K

Quick calculation: 405 queries save 15 mins per query = 6075 minutes = 101 hours at $60/hr (if not more) = **$6075** saved
How is it Used?

Target Suitability - “Spending Smart”

POSSIBLE REMEDIES

- Go back to linguist for more editing
- Alert of higher LQA risk
- Use information to retrain MT engine (dynamic?)
- Map to client LQA methodology
- Spend LQA $$ where it counts
- Confirm MTQE
- Confirm PE Distance and/or TER
- Confirm productivity metrics
How is it Used?

**Summary View**

- How many features failed?
- Pass/Fail/Review per segment
- Aggregated to pass/fail per file
How is it Used?

Summary View

- Passes/fails per domain
- Passes/fails per locale pair

2.
How is it Used?
Garbage In, Garbage Out

- TRACING SOURCE TO TARGET CORRELATIONS
- POOR SOURCE LEADS TO POOR TARGET

<table>
<thead>
<tr>
<th>File name</th>
<th>ADJ Count</th>
<th>NOUN Count</th>
<th>PROPN Count</th>
<th>Word Count</th>
<th>Long Word Count</th>
<th>Complex Word Count</th>
<th>Nominalization Count</th>
<th>LM Score</th>
<th>FleschReadingEase</th>
<th>ADJ Count</th>
<th>NOUN Count</th>
<th>PROPN Count</th>
<th>Word Count</th>
<th>Long Word Count</th>
<th>Complex Word Count</th>
<th>Nominalization Count</th>
<th>LM Score</th>
<th>FleschReadingEase</th>
</tr>
</thead>
<tbody>
<tr>
<td>TASK10195629</td>
<td>0.6666667</td>
<td>3.333333</td>
<td>1.1111111</td>
<td>9.266667</td>
<td>3.555556</td>
<td>0.557778</td>
<td>0.35555556</td>
<td>64.8734</td>
<td>53.9952</td>
<td>0.8888888</td>
<td>2.333333</td>
<td>1.977778</td>
<td>9.088889</td>
<td>3.844444</td>
<td>0.911111</td>
<td>0.2666666666666666</td>
<td>56.1765</td>
<td>55.52887</td>
</tr>
<tr>
<td>TASK10195633</td>
<td>0.9545455</td>
<td>4.318182</td>
<td>0.681818</td>
<td>12.454555</td>
<td>5.363636</td>
<td>3.181818</td>
<td>0.454545455</td>
<td>257.5985</td>
<td>36.08856</td>
<td>1.727273</td>
<td>3.363636</td>
<td>1.363636</td>
<td>12.045454</td>
<td>6.136364</td>
<td>2.272727</td>
<td>0.2772727272727272</td>
<td>372.648</td>
<td>59.00989</td>
</tr>
<tr>
<td>TASK10195637</td>
<td>0.766667</td>
<td>3.266667</td>
<td>1.366667</td>
<td>9.866667</td>
<td>3.633333</td>
<td>1.633333</td>
<td>0.3</td>
<td>411.9258</td>
<td>55.52377</td>
<td>1.121212</td>
<td>2.727273</td>
<td>2.303030</td>
<td>9.242424</td>
<td>3.878788</td>
<td>0.969997</td>
<td>0.2121212121212121</td>
<td>2095.75</td>
<td>58.82482</td>
</tr>
<tr>
<td>TASK10227020</td>
<td>1.333333</td>
<td>3.966102</td>
<td>0.711064</td>
<td>14.525424</td>
<td>2.694915</td>
<td>1</td>
<td>0.355932203</td>
<td>445.9728</td>
<td>52.33312</td>
<td>1.542373</td>
<td>3.923203</td>
<td>0.881356</td>
<td>15.45763</td>
<td>5.745763</td>
<td>1.474957</td>
<td>0.4406779666666666</td>
<td>607.4004</td>
<td>55.07588</td>
</tr>
<tr>
<td>TASK10299494</td>
<td>1.141587</td>
<td>3.975169</td>
<td>0.619048</td>
<td>12.426576</td>
<td>4.380952</td>
<td>2.338095</td>
<td>0.438514293</td>
<td>1075.118</td>
<td>50.6495</td>
<td>1.452381</td>
<td>2.142386</td>
<td>1.309524</td>
<td>12.28371</td>
<td>4.785714</td>
<td>1.047619</td>
<td>0.357142857142857</td>
<td>793.157</td>
<td>58.67438</td>
</tr>
<tr>
<td>TASK10299496</td>
<td>2.433333</td>
<td>8.266667</td>
<td>1 23.83333</td>
<td>9.266667</td>
<td>6.166667</td>
<td>1</td>
<td>0.833333333</td>
<td>227.824</td>
<td>29.01318</td>
<td>2.266667</td>
<td>7.333333</td>
<td>1.366667</td>
<td>10.566666</td>
<td>2.866667</td>
<td>0.911111</td>
<td>0.2666666666666666</td>
<td>56.1765</td>
<td>55.52887</td>
</tr>
<tr>
<td>TASK10354283</td>
<td>0.608696</td>
<td>2.717391</td>
<td>0.739131</td>
<td>6.902174</td>
<td>2.706522</td>
<td>1.141304</td>
<td>0.293478261</td>
<td>268.863</td>
<td>42.92559</td>
<td>0.684783</td>
<td>2.26087</td>
<td>1.336957</td>
<td>7.141304</td>
<td>3.108696</td>
<td>0.652174</td>
<td>0.217391304040404</td>
<td>1856.129</td>
<td>58.27621</td>
</tr>
</tbody>
</table>
How is it Used?

How Bad is the File?

More than half of the file has 6 or more nouns
Half of the file has 8 long words or more
How is it Used?

A Telling Example

Today’s machines enable industrial workers to carry out complex Computer Aided Design, Manufacturing and Engineering (CAD, CAM, CAE) operations, model Computational Fluid Dynamics (CFD), accomplish thermal, stress and fatigue analysis, or visualise and test designs and models using immersive Virtual Reality (VR).

And now the statistics

- 42 words
- 22 nouns
- 19 long words
- 9 complex words

List of nouns

Today | machines | workers | Computer | Design | Manufacturing | Engineering | CAD | CAM | CAE | operations | model | Computational | Fluid | Dynamics | CFD | thermal | stress | fatigue | analysis | designs | models
How is it Used? Under the **Hood**

NLP frameworks
Human validation
Predictive modeling

**NLP**
- Language Models
- Tokenization & PoS tagging (NLP frameworks)
- Readability features (FleschKincaid, complex words, long words, nominalization)

**Human Input**
Definition of features, review and calibration of features, fine-tuning, data analysis

**Predictive Modeling**
Long-term vision if sufficient data available
How is it Used?

Process Optimization

Reducing time to market and costs while improving linguist acquisition and retention

- LQA Time Saved: 15-20%
- LQA Pass Rate Improvement: 20%
- LQA Spend Reduction: 10%
Continued human validation

Build predictive models using machine learning (ML) algorithms

Human validation comment

“I think this is a very interesting tool that has a lot of potential. The output statistics provide some interesting insights about the nature and style of the source, and more importantly, also the target text. With the help of these figures, a source text can be analyzed for its complexity, while a translation can be characterized and possibly rated with regard to certain stylistic guidelines.”
Thank you