Linguistic QA for MT of User-Generated Content at eBay

eBay Machine Translation Linguistic Team
October 2014
1. MT @ eBay: introduction
2. Role of Machine Translation Language Specialist
3. LQA: Challenges and Solutions
4. Q&A
MT @ EBAY: INTRODUCTION
Why we do it: promoting cross-border trade

Geo Expansion: localized interface + MT

August 2013 – in-house SMT

January 2014 – MT for Russian
eBay MT translates:

- search queries
- item titles
eMT currently supports the following language pairs:

<table>
<thead>
<tr>
<th>Language</th>
<th>Search Queries</th>
<th>Item Titles</th>
<th>Default?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Russian</td>
<td>RU&gt;EN</td>
<td>EN&gt;RU</td>
<td>Y</td>
</tr>
<tr>
<td>Spanish (Lat.Am. &amp; Spain)</td>
<td>ES&gt;EN</td>
<td>EN&gt;ES</td>
<td>Y</td>
</tr>
<tr>
<td>Br. Portuguese</td>
<td>PT&gt;EN</td>
<td>EN&gt;PT</td>
<td>Y</td>
</tr>
<tr>
<td>French</td>
<td>FR&gt;EN</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>Italian</td>
<td>IT&gt;EN</td>
<td></td>
<td>Y</td>
</tr>
</tbody>
</table>
Challenges for MT at eBay
Challenges for MT at eBay

1. Variety of context:
- ~4K categories on ebay.com
Challenges for MT at eBay

2. User-generated content:

- Spelling errors/typos/mixed language
  
sansung samsug samsumg samung amsung samnsung smsung samsuns

- Keylock issues
  
ω3ρψτυ 5 Ьl

- Syntax
  
Oval Picture Photo Frame Tibet Silver Golden Copper Bronze Pendant Charm Finding

- Improper, broken English
  
ull buy em rii nah thru paypal
  
you will buy them right now through PayPal
To improve performance of SMT:

- improve quality of training corpora
- improve tokenization algorithms
- improve named-entity recognition
ROLE OF MTLS

Machine Translation Language Specialists
Role of MTLS:

- Evaluate and analyze raw MT output
- Review and evaluate MT output translated and/or post-edited by vendor
- Ongoing post-launch LQA on SQ and ITT in production
- Named-entity/keylock tagging

Ensuring and maintaining high quality of training and testing corpora
Working with Vendors:

Training data > post-edited (>300,000 words per language per training cycle)

- Search queries
- Item titles
- Item descriptions

Testing data: > human translation (> 40,000 words per language per training cycle)

- Search queries
- Item titles
- Item descriptions
Basic LQA Flow

Training data:
- Raw MT output
- Vendor postedited output
- MTLS review
- Data fed into the engine

Testing data:
- Source text
- Vendor translates
- MTLS review
- Data used for reference
LQA: CHALLENGES AND SOLUTIONS

Vendors and MTLS
Challenges for Vendor

• Unfamiliar multi-domain content (~4K categories, ~30 meta categories)

• Lack of context

• User-generated source content: typos, broken word order, inconsistent terminology usage

• eBay-specific approach and requirements aimed at an MT engine, not human readers (preserve the syntax of IT, avoid preferential stylistic changes, etc.)
Our Solution

- Compiled set of modular guidelines with main part applicable to most projects
- Provide reference materials: mindmaps, glossaries (brand names, acronyms, eBay-specific terms), training videos
- Provide context by adding category information
- Limit research time to 2-3 min max and recommend staying within eBay
- Allow 5% of strings to be left unedited if incomprehensible
- Provide extensive training and ongoing communication and support
Visual Guidelines (mindmaps)
Challenges for MTLS

Review Requirements:

- Time-efficiency: large volumes, tight deadlines
- Effective sampling: consistency of quality throughout the file
- Systematic feedback: easily implemented, serve as additional training
- Error metrics: reflect eBay-specific requirements
Our Solution

- **Feedback and Error Metrics:** we customized error typology

**Challenging areas:**
- Meaning vs spelling
- Fluency vs style
- Localizing
- Word order
- Brand names
- Acronyms

### Issue categories

<table>
<thead>
<tr>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy - Mistranslation</td>
</tr>
<tr>
<td>Accuracy - Addition/Omission</td>
</tr>
<tr>
<td>Accuracy - Untranslated words</td>
</tr>
<tr>
<td>Accuracy - DNT word was translated</td>
</tr>
<tr>
<td>Accuracy - MT was not postedited</td>
</tr>
<tr>
<td>Accuracy - Terminology</td>
</tr>
<tr>
<td>Language - Spelling</td>
</tr>
<tr>
<td>Language - Grammar</td>
</tr>
<tr>
<td>Language - Punctuation</td>
</tr>
<tr>
<td>Language - Capitalization</td>
</tr>
<tr>
<td>Style Guide - Do not localize</td>
</tr>
<tr>
<td>Style Guide - Source not corrected</td>
</tr>
<tr>
<td>Style Guide - Acronyms</td>
</tr>
<tr>
<td>Style Guide - Other</td>
</tr>
<tr>
<td>Feedback not implemented</td>
</tr>
</tbody>
</table>

### Severities

- Minor
- Major
- Critical
- Neutral
Our Solution

Semi-Automated **3-Step Sampled Review Process:**

1. **Beyond Compare:** file comparison

2. **Checkmate:** predefined and custom checks on bilingual documents. Error reports to share with vendors.
   - **LanguageTool:** grammar and spelling checker; integrated with Checkmate; predefined and custom rules.

3. **Manual Review:** using internally developed template
Benefits:

- Scalability: documentation and QA process easily applicable to new languages
- Productivity/time efficiency: throughput improved x2
- Reliability: combination of tools and manual review
- Preciseness: customized and adapted to each language