Getting Started With Machine Translation (MT 101)

Dion Wiggins
Chief Executive Officer
dion.wiggins@asiaonline.net
The MT Situation Today

**MT is Pervasive and Easily Available**

- **Free Online MT Systems**
  - Google Translate
  - Bing Translate
  - 250 – 500 Million Users/Day

- **Open Source**
  - Moses / Apertium

- **Instant Custom MT**
  - Upload & Pray

- **Expert MT Systems**

Copyright © 2015, Asia Online Pte Ltd
Different Base MT Technologies

Rule-Based (RBMT) or Transfer-Based
- Systran, ProMT, Lucy Software, Lionbridge, SDL, Toshiba, AppTek

Statistical (SMT) or “Data-driven”
- Microsoft, Google, IBM, BBN, SDL BeGlobal, Asia Online Language Studio
- Many smaller open source based initiatives.
- Moses Open Source SMT
2 Types of Hybrid MT

Language Studio™ Hybrid Rules and SMT Engine Model

**Pre-Processing Rules**
- Open Process
- Fully Customizable
- Complete Control
- No Limits

Rules prepare the source text for better translation and guide the SMT process

**Statistical Machine Translation**
- Open Process
- Fully Customizable
- Complete Control
- No Limits

**Post-Processing Rules**
- Open Process
- Fully Customizable
- Complete Control
- No Limits

Rules fine tune the SMT output and offer additional flexibility

Competitors “Hybrid” Rules and Corrective Statistical Engine Model

**Translation Rules**
- Black Box
- Where “Magic” Happens
- Limited Control
  - Dictionaries Only

**Statistical Correction of Rules Errors**
- Black Box
- Where “Magic” Happens
- Limited Control
  - Phrases Only

Get it right the first time!!
Why treat the symptoms when you can treat the cause?

Band-aid to fix a broken solution
### Pre-Processing Rules
- Sentence Segmentation
- Word Segmentation
- Phrase Reordering
- Dates and Numbers
- Patterns, Formulas etc.
- Pre-Normalization
- Spell Checking
- Custom Runtime Glossary
- Pre-Formatting

### Post-Processing Rules
- Capitalization
- Post-Formatting
- Grammar Checking
- Post-Normalization
- XML Tag Reinsertion
- Currency Conversion
- Cross Referencing
- Other custom post processing

### Statistical Machine Translation

Pre-Processing and Post-Processing Rules are implemented in Java, which can be customized to fit almost any requirements. This is a true hybrid approach, where rules are used to guide the statistical engine to greater accuracy.
The Many Variants of MT

- **Most DIY MT**
  - Low Quality of Translation Output

- **Generic MT**
  - Spanish, Portuguese, French
  - Hungarian, Turkish, Korean, Hindi, Zulu, Swahili

- **Successful DIY MT**
  - Expert
  - Proprietary

Copyright © 2015, Asia Online Pte Ltd
## MT Quality / Productivity Relationship

<table>
<thead>
<tr>
<th>MT System Quality</th>
<th>Productivity Implications of Better Quality MT Engines</th>
</tr>
</thead>
</table>
| Free Online Engines       | Can be useful in some languages but often lower productivity than using TM alone and impossible to adapt to specific needs  
1,000 to 3,000 Words/Day per human editor  
*Average segment quality = ~15 to 40% TM Fuzzy Match* |
| Human TEP Process         | Typically produce 2,500 Words/Day per Translator       |
| Low Quality - Moses       | Less than 5% of these systems can outperform free online MT and best case productivity may be in the 3,000 Words/Day range  
*Average segment quality = 50% - 60% TM Fuzzy Match* |
| Average Expert System     | These systems can provide 5,000 to 7,000 Words/Day per editor  
*Average segment quality = 60% - 75% TM Fuzzy Match* |
| Superior Expert           | These systems can provide 9,000 to 12,000 Words/Day per editor  
*Average segment quality = 70% - 85% TM Fuzzy Match* |
| Exceptional MT            | These systems can provide 12,000+ Words/Day per editor  
*Average segment quality = 80% - 90% TM Fuzzy Match* |
Key Concepts and Approaches
On MT Customization
What is MT Customization?

Adapting an MT engine to **minimize** the amount of post editing required

Adapting an MT engine so that it translates more in **context**

= The Highest Possible MT Output Quality to get work done **Faster** and at **Lower Cost**
There are many factors that impact machine translation output quality of a custom engine. These include:

- The complexity of the language pair and complexity of the subject domain
- The quality of the source content
- The amount and quality of translation memories available for training
- The amount and quality of bilingual dictionaries and glossaries available for training
- The domain suitability of the translation memories
- The amount of effort put in to the initial customization for normalization and rule refinement
- The amount of effort put in by the customer to work with Language Studio™ Linguist to identify and resolve issues
- The amount of post edited data fed back into the engine for incremental quality improvement
Skills and Knowledge Needed for Creating High Quality Custom MT Engines

<table>
<thead>
<tr>
<th>Data Preparation and Selection</th>
<th>Translation Quality Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>• What is the right data to upload for my MT system?</td>
<td>• All the same questions as when initially preparing and selecting data</td>
</tr>
<tr>
<td>• How should I prepare my data?</td>
<td>• How can I know the cause of the error?</td>
</tr>
<tr>
<td>• What cleaning can I do that the DIY vendor does not do?</td>
<td>• How can I fix formatting errors?</td>
</tr>
<tr>
<td>• What impact will my data have on the MT system?</td>
<td>• How can I fix word order issues?</td>
</tr>
<tr>
<td>• Will the data I upload improve or decrease translation quality?</td>
<td>• How can I fix incorrect terminology errors?</td>
</tr>
<tr>
<td>• What will mixing data from multiple domains do to my MT system?</td>
<td>• How can I fix inconsistent terminology?</td>
</tr>
<tr>
<td>• Should I add some or all of the TAUS data to my system?</td>
<td>• How can I fine tune so that non-breaking spaces are between numbers and units?</td>
</tr>
<tr>
<td>• Should I upload third party data?</td>
<td>• Should I remove some of the data I previously uploaded?</td>
</tr>
<tr>
<td>• Does the DIY vendor already include the same third party data?</td>
<td>• How do I measure quality accurately?</td>
</tr>
<tr>
<td>• Should I mix data from customers in one engine or keep them separate?</td>
<td>• What will the different quality metrics tell me and which are more reliable?</td>
</tr>
<tr>
<td>• How can I normalize my terminology so that terms are consistent?</td>
<td>• Should I rely on human quality metrics or automated quality metrics?</td>
</tr>
<tr>
<td>• How can I get consistent writing style?</td>
<td>• How often should I retrain my custom MT engine?</td>
</tr>
<tr>
<td></td>
<td>• What can I do without retraining the engine to improve quality?</td>
</tr>
<tr>
<td></td>
<td>• How much data do I need to improve quality?</td>
</tr>
</tbody>
</table>

OR...
Language Studio™ Technology Reduces the Effort to Quality Ratio

Asia Online “Your MT Department” Effort
- Majority of all Technical Tasks
- Custom Engine Development Plan
- Quality Improvement Plan
- Issue Analysis and Resolution
- Terminology Normalization
- Unknown Word Detection and Resolution
- Gap Analysis
- Data Manufacturing
- Rule Creation
- Translation Quality Analysis
- Many more Quality Centric Tasks...

Your Effort
- Term Management
- Engine Definition
- Evaluations

MT Effort
- DIY Upload TM
- Click to Translate

<20-40% Productivity Gain

Google
DIY

150-300% Productivity Gain

Translation Quality
The Importance of Domain Focus

<table>
<thead>
<tr>
<th>Language Pair</th>
<th>Top-Level Domain</th>
<th>Engines/Sub-Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>EN-ES</td>
<td>Automotive</td>
<td>Honda</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cars</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Motorbikes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Toyota</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cars</td>
</tr>
<tr>
<td></td>
<td></td>
<td>User Manuals</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Engineering Service Manuals</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Marketing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Service Reports</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Customization Level:</th>
<th>Generic</th>
<th>Domain</th>
<th>Client</th>
<th>Product</th>
<th>Target Audience / Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Productivity Gain:</td>
<td>???</td>
<td>&lt; 20-40%</td>
<td>50%+</td>
<td>90%+</td>
<td>150-300%+</td>
</tr>
</tbody>
</table>

- Generic MT from Google, Bing, etc. offers unknown productivity gains and often productivity loss due to lack of control.
- Instant MT offer < 20-40% productivity gains due to top domain only focus and “dirty data SMT” customization model.
- Language Studio™:
  - Targets of **150-300%+** productivity gains with granular sub-domain “clean data SMT” approach.
  - Provides complete control of writing style, terminology and is mapped to target audience reducing editing effort.
<table>
<thead>
<tr>
<th>Spanish Original Before Translation:</th>
<th>Se necesitó una gran maniobra política muy prudente a fin de facilitar una cita de los dos enemigos históricos.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business News After Translation:</td>
<td>Significant amounts of cautious political maneuvering were required in order to facilitate a rendezvous between the two bitter historical opponents.</td>
</tr>
<tr>
<td>Spanish Original Before Translation:</td>
<td>Se necesitó una gran maniobra política muy prudente a fin de facilitar una cita de los dos enemigos históricos.</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Business News After Translation:</td>
<td>Significant amounts of cautious political maneuvering were required in order to facilitate a rendezvous between the two bitter historical opponents.</td>
</tr>
<tr>
<td>Children’s Books After Translation:</td>
<td>A lot of care was taken to not upset others when organizing the meeting between the two long time enemies.</td>
</tr>
</tbody>
</table>

Language Studio is the only platform to deliver this level of stylistic control.
Unique Customization Plan

- No Data
- A Little Data
- Sufficient Data
- Large Amounts of Data
Customization is a Process, Not an Upload

- Data Preparation
  - Data Cleaning
- Data Collections
- Translate
- Diagnostics and Fine Tuning
- Training
- Quality Assurance
- Language Pair Foundation Data
- Domain Foundation Data
- Original Translation Sources
4 Data Elements of Customization

**Sub-Domain Specific Data**

- Client Data
- Manufactured Data

**Asia Online Foundation Data**

- Language Pair Foundation
- Domain Foundation

= Custom Engine
Source Language Data

**Source Material To Be Translated**
Source material can be analyzed and processed to further improve quality.

**URLs and Key Words for Data Manufacturing**
URLs and keywords that will help to locate content that represents data to be translated (Only needed for language pairs where limited data is available)

Bilingual Data

**Bilingual Translation Memories**
In domain historical translations in source and target language.

**Bilingual Dictionaries and Glossaries**
In domain and client specific glossaries and dictionaries.

**Source Language Non-Translatable Terms**
Source language terms such as product names and brand names that should not be translated.

Target Language Data

**Target Language Monolingual Data**
Monolingual target language text and URLs of in-domain websites.

**Style Guides**
Rules can be added to match client style guide requirements.

**URLs and Key Words for Data Manufacturing**
URLs and keywords that will help to located content that represents target language style and grammar (Only needed for language pairs where limited data is available)
Understanding Foundation Data

A high quality engine is as much about what data is not in the engine as it is about what data is in the engine

- High quality data from general domains that is used as a base to build upon
- Data has had extensive cleaning and validation
- Data is from trusted sources
- Data is balanced and normalized
- Data that might negatively influence an engine is removed
High quality in domain data should be the primary data that is used to build statistical models.

**Dirty Data SMT Model**
- **Data**
  - Gathered from *as many sources as possible*.
  - Domain of knowledge does not matter.
  - Data quality is not important.
  - Data quantity is important.
- **Theory**
  - Good data will be more statistically relevant.

**Clean Data SMT Model**
- **Data**
  - Gathered from *a small number of trusted quality sources*.
  - Domain of knowledge must match target.
  - Data quality is very important.
  - Data quantity is less important.
- **Theory**
  - Bad or undesirable patterns cannot be learned if they don’t exist in the data.
Advanced Data Manufacturing

- Gap analysis
- Unknown word analysis
- Inflected form creation
- Grammatical structure creation
- Syntactic clause creation
- Terminology normalization
- Automated rule generation
- Over 150 tools and processes
Quality requires an understanding of the data

There is no exception to this rule
Post-Editing MT
I may not be perfect, but parts of me are excellent!
Adding MT to your Workflow

Translation Memory (TM) Match

- Matched Sentences
- Non-Matched Sentences

Machine Translate

- Correct TM Match
- Human Translate
- Correct MT Output

Translate via your preferred TMS

1. TM match. A high fuzzy score (usually 85%) is used as raw MT will often be better than a lower scoring fuzzy match.
2. Send unmatched segments to MT.
3. MT Output can be corrected by a translator or in the case of high quality MT, often bypass this step and go directly to an editor.

Edit and Proof using your preferred editing tools.

1. Different projects have different budgets and quality requirements.
2. The high quality segments for a given project should be sent to Asia Online to improve the quality of future translations. This is usually just prior to publication.
Pro Tip – Focus on a Single Task

1. Follow the 3 seconds rule
2. Should it be edited or translated?
3. If it takes you longer than 3 seconds to decide, then it should probably be translated.

Translate Later
You or another translator can translate the lower quality segments.

Edit Now
Keep in the “post-edit” mind set.

Evidence from customers taking this approach indicates that can deliver as much as 50% performance difference and increases consistency.
3 Typical Stages of TEP Process

- Tried and proven method / approach
- Well understood by both LSP industry and end clients
• Same person is responsible for both post editing and re-translating non-editable segments
• There is no edit stage after the re-translate, so a key quality check is missing
• Simply replacing the Human Translation with Machine Translation is insufficient
Correct Approach to Workflow

Translate

Machine Translation

Human Translation

Edit

Quality Check
Is MT Segment
Editable?

Yes

Post Edit

No

Edit

Proof

Note: Human Translator and Post Editor can be the same person, but Human Translator and Editor must not be the same person.
The Language Studio™ 4 Step Quality Plan

1. Customize
Create a new custom engine using foundation data and your own language assets.

2. Measure
Measure the quality of the engine for rating and future improvement comparisons.

3. Improve
Provide corrective feedback removing potential for translation errors.

4. Manage
Manage translation projects while generating corrective data for quality improvement.

Quality
The quick brown fox jumped over the lazy dog

Copyright © 2015, Asia Online Pte Ltd
How to Improve Quality

Additional Training Data

- Each custom engine is a living engine and constantly improves with use. There are many sources of data that can improve an engine’s translation quality.

- **Posted Edited Machine Translations**
  Post editing of raw MT rapidly improves translation quality.

- **Data Manufacturing**
  Language Studio™ will analyze edits and other data and manufacture new data to improve quality.

- **Bilingual Translation Memories**
  Additional in domain historical translations in source and target language that were not included in earlier training.

- **Target Language Monolingual Data**
  Additional monolingual target language text and URLs of in-domain websites that were not included in earlier training.

- **Bilingual Dictionaries and Glossaries**
  Additional in domain and client specific glossaries and dictionaries that were not included in earlier training.

- **Source Language Non-Translatable Terms**
  Additional source language terms that should not be translated that were not included in earlier training.

Runtime Improvements

- Fine tuning to specific formats and style guide requirements can be performed at runtime without retraining the engine.

- Pre-Translation JavaScript
- Pre-Translation Corrections
- Non-Translatable Terms
- Runtime Glossary
- Post-Translation Adjustments
- Post-Translation JavaScript

These features enable:

- Normalization of terms
- Control of preferred terminology
- Mapping of complex rules as specified in the style guide

Based on data available, Asia Online will design a quality improvement plan to provide the optimal path to higher quality translations. All at no extra cost!!!
Customizing Translations at Runtime

**Before Machine Translation**

*Source text is processed and modified.*

**Pre-Translation JavaScript (JS)**
- Complex pre-processing can be customized via JavaScript.

**Pre-Translation Corrections (PTC)**
- A list of terms that adjust the source text fixing common issues and making it more suitable for translation.

**Non-Translatable Terms (NTT)**
- A list of monolingual terms that are used to ensure key terms are not translated.

**Runtime Glossary (GLO)**
- A list of bilingual terms that are used to ensure terminology is translated a specific way.

**After Machine Translation**

*Target text is processed and modified.*

**Post Translation Adjustment (PTA)**
- A list of terms in the target language that modify the translated output. This is very useful for normalization of target terms.

**Post Translation JavaScript (JS)**
- Complex post-processing can be customized via JavaScript.

Runtime customizations can be applied in 2 forms:

**Default:** Applied to all jobs.

**Job Specific:** A different set of customizations can be applied for different clients.
Building PEMT Leverage

- **Learn How to Give Corrective Feedback to MT**
  - Better MT Output = Higher Productivity = Higher Pay
  - Learn to communicate with MT developers

- **Understand Error Patterns**
  - Provide corrective strategies

- **Learn to Use Power Tools To Improve Productivity**
  - Some MT solutions have run time tools to improve output
  - Add Glossaries, REGEX Rules, DNT Lists

- **Improve and Clean Before Editing**
  - Fix Dumb Errors early and quickly to minimize repetition
  - Use Run Time Tools
  - Retrain the engine frequently so edits get easier

- **Integrate Into Your Preferred Editing Tools**
  - Can be as simple as importing TMX files
Integrated with 30+ Third Party Systems
Pre-Analysis of Data To Be Translated

If you know what you intend to translate, quality can be further assured by analyzing the text.

• High Frequency Terms
  – All text to be translated is analyzed for frequency of occurrence
  – High frequency terms are machine translated and then verified by human
  – Term length ranges from 1 word to 8 words

• People, Place Names, Facility Names, Locations
  – Often mistranslated if not pre-defined.
  – Extract and determine how they should be translated or left in their original form.
**Just a short walk from Times Square**

The following n-grams, or word strings, can be generated from this sentence.

<table>
<thead>
<tr>
<th>1 Gram</th>
<th>2 Grams</th>
<th>3 Grams</th>
<th>4 Grams</th>
<th>5 Grams</th>
</tr>
</thead>
<tbody>
<tr>
<td>Just a short walk from times square</td>
<td>just a short walk from times square</td>
<td>just a short walk from times square</td>
<td>just a short walk from times square</td>
<td>just a short walk from times square</td>
</tr>
</tbody>
</table>

**Identify high frequency phrases (ngrams) to translate**

- first gulf
- ghadeer at
- holiday apartment
- insurance program
- investors might
- marlin beach
- nar pulse

Copyright © 2015, Asia Online Pte Ltd
In the real world, a term can have a variety of possible translations. Statistically a term can have even more.

Trained Bilingual Data

List of possible output based on statistics

Database
Relational Database
Relational Database Management System
DB
DBMS
RDBMS....

Client Preferred Terminology

base de datos relacional
Terminology Consistency

*MT + Human can deliver higher quality output than human only translation processes*

- Language Studio™ provides tools and processes for normalization of terminology
- Benefits include cost reductions, faster deliverables, higher customer satisfaction and happier post editors

<table>
<thead>
<tr>
<th>Source</th>
<th>Human Reference</th>
<th>Language Studio™ Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Install</em> a new gasket and the inlet housing with the 3 bolts.</td>
<td>Coloque una junta nueva y el alojamiento de entrada con los 3 pernos.</td>
<td>Instale una junta nueva y el alojamiento de entrada con los 3 pernos.</td>
</tr>
<tr>
<td><em>Install</em> the brush holder to the generator coil with the 2 screws.</td>
<td>Instale el portaescobillas en la bobina del generador con los 2 tornillos.</td>
<td>Instale el portaescobillas en la bobina del generador con los 2 tornillos.</td>
</tr>
<tr>
<td><em>Remove</em> the bolt and hole cover from the transaxle.</td>
<td>Quite el perno y la cubierta del orificio de la caja del transeje.</td>
<td>Extraiga el perno y la cubierta del orificio del transeje.</td>
</tr>
<tr>
<td><em>Remove</em> the clamps and connectors, and then disconnect the wire harness.</td>
<td>Extraiga las abrazaderas y los conectores y, a continuación, desconecte el mazo de cables.</td>
<td>Extraiga las abrazaderas y los conectores, y a continuación desconecte el mazo de cables.</td>
</tr>
</tbody>
</table>
Time to Achieve ROI
Understand Where You Are Starting From

If you don’t know where you are now, how do you know if you are getting better or worse?

• Start with a clear understanding of current human levels of productivity and quality
• Measure productivity levels on each notable project to determine if there is progress
• Focus on production productivity, not the number of translation errors
Common Mistake: Measure quality of initial engine and determine costs from this engine.

- Many issues in data and formatting can only be seen once an engine has been customized.
- One of the most important metrics is how quickly a custom engine can improve. This differs between vendors, engines, language pair and domain.
- Many of the most significant issues can be addressed very quickly.
- Metrics should begin once and initial correction/diagnostic phase has been completed and corrective action taken.
1. **Customize**
Create a new custom engine using foundation data and your own language assets.

2. **Measure**
Measure the quality of the engine for rating and future improvement comparisons.

3. **Improve**
Provide corrective feedback removing potential for translation errors.

4. **Manage**
Manage translation projects while generating corrective data for quality improvement.
Quality Evolution via Corrective Feedback

Key
- Green: Correct Raw MT
- Red: Mistranslation
- Orange: Syntax/Grammar
- Maroon: Terminology
- Brown: Spelling
- Brown: Punctuation

Initial System
- Correct
- Correct
- Correct

Spelling and Terminology
- Correct
- Correct
- Correct

Targeted Corrections of Bad Learning
- Correct
- Correct
- Correct

Human Feedback
- Correct
- Correct
- Correct

Human Feedback can raise the raw output to previously unseen quality levels
Language Studio™ uses the Clean Data SMT approach with expert human guided granular sub-domains.

**Comparison of Data Requirements for Improvement with Dirty Data and Clean Data SMT Models**

- **DIY Vendor Dirty Data SMT Baseline**
  - Initial Customization
  - Improvement
  - 20% Required for Noticeable Improvement

- **Language Studio™ Clean Data SMT Foundation**
  - Initial Customization
  - Improvement

- **Language Studio™ Clean Data SMT Foundation**
  - Initial Customization
  - Improvement
  - < 0.1%

Copyright © 2015, Asia Online Pte Ltd
High Value Skills for MT Success

**Shift from TM and Translation Segment Match Worldview to a Corpus Orientation**

- Data Identification and Acquisition Skills
- Rapid Quality Assessment of MT
- Skills with Run-Time Tools and Processes
- Clear understanding of benchmarks
- Post-Editor Training Capabilities
- Web resources of critical data
- Identify best post-editors
- Continuous productivity measurement
- Re-train frequently to raise quality
- Think in error patterns not just individual errors
Results of Refinements

GOAL: Progressively develop engine quality to a level that exceeds the equivalent productivity of an 85% translation memory fuzzy match.

Post Editing Cost
- MT learns from post editing feedback and quality of translation constantly improves.
- Cost of post editing progressively reduces as MT quality increases after each engine learning iteration.

Post Editing Effort Reduces Over Time
- The post editing and cleanup effort gets easier as the MT engine improves.
- Initial efforts should focus on error analysis and correction of a representative sample data set.
- Each successive project should get easier and more efficient.
• Directly linked to time to improve quality
• Faster improvement delivers faster ROI
• Many customers recover costs on their first project
• Projects progressively become lower cost
Understand the Human Factors

- Manage Expectations of Key Players
  - PMs, Editors, Translators
  - Clients

- Provide Many Editing Examples

- Get MT Output to Acceptable Levels
  - Fix Dumb Errors early and quickly to minimize repetition
  - Cycle quickly through initial MT development cycles
  - Take feedback seriously and incorporate quickly
  - Retrain the engine frequently so edits get easier

- Establish Fair & Reasonable Compensation
  - Based on benchmarks and real throughput
  - Err on the side of overcompensation than under

- Community Management
  - Train Editors
  - Communicate & Collaborate
  - Quality Management
What is your BLEU score?

This is the single most irrelevant question relating to translation quality, yet one of the most frequently asked.
The Basics of BLEU Scores

**Original Source:**
- The original sentences that are to be translated.

**Human Reference:**
- The gold standard of what a high quality human translation would look like.

**Translation Candidate:**
- This is the translated output from the machine translation system that you are comparing.

**Note:** Multiple machine translation candidates can be scored at one time to compare against each other. E.g. Asia Online, Google, Systran

**3 Measurement Tools**
- Human Quality Assessment
- Automated Quality Metrics
- Sentence Evaluation

Copyright © 2015, Asia Online Pte Ltd
### Evaluating 1,000 Segment Test Set

#### Automated Metrics

<table>
<thead>
<tr>
<th></th>
<th>Language Studio™</th>
<th>Google</th>
<th>Bing</th>
<th>Systran</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLEU Case Sensitive</td>
<td>71.86</td>
<td>40.21</td>
<td>35.71</td>
<td>26.67</td>
</tr>
<tr>
<td>BLEU Case Insensitive</td>
<td>76.71</td>
<td>41.27</td>
<td>36.73</td>
<td>27.52</td>
</tr>
<tr>
<td>F-Measure Case Sensitive</td>
<td>86.52</td>
<td>69.98</td>
<td>68.26</td>
<td>60.67</td>
</tr>
<tr>
<td>F-Measure Case Insensitive</td>
<td>88.93</td>
<td>70.62</td>
<td>69.06</td>
<td>61.06</td>
</tr>
<tr>
<td>TER Case Sensitive</td>
<td>21.75</td>
<td>44.08</td>
<td>46.36</td>
<td>55.30</td>
</tr>
<tr>
<td>TER Case Insensitive</td>
<td>19.93</td>
<td>43.81</td>
<td>46.00</td>
<td>55.18</td>
</tr>
</tbody>
</table>

#### Human Review

<table>
<thead>
<tr>
<th></th>
<th>Language Studio™</th>
<th>Google</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perfect, Exact Match to Reference</td>
<td>369</td>
<td>8</td>
</tr>
<tr>
<td>Perfect, But Different to Reference</td>
<td>310</td>
<td>270</td>
</tr>
<tr>
<td>Perfect and Better than Reference</td>
<td>30</td>
<td>-</td>
</tr>
<tr>
<td>Minor Edits Required (1-2 Edits)</td>
<td>230</td>
<td>111</td>
</tr>
<tr>
<td>Medium Edits Required (3-4 Edits)</td>
<td>61</td>
<td>291</td>
</tr>
<tr>
<td>Bad – Needs Retranslation</td>
<td>-</td>
<td>320</td>
</tr>
<tr>
<td></td>
<td>1,000</td>
<td>1,000</td>
</tr>
</tbody>
</table>

61% Needed considerable work
## Cost / Savings Analysis

<table>
<thead>
<tr>
<th>Word Costs</th>
<th>Human</th>
<th>Language Studio™ MT + Human</th>
<th>Google MT + Human</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translation</td>
<td>0.1700</td>
<td>0.0052</td>
<td>0.00014</td>
</tr>
<tr>
<td>Post Editing</td>
<td>0.0500</td>
<td>0.0500</td>
<td>0.1300</td>
</tr>
<tr>
<td>Proof Reading</td>
<td>0.0300</td>
<td>0.0250</td>
<td>0.0300</td>
</tr>
<tr>
<td>Project Size (Words)</td>
<td></td>
<td>1,000,000</td>
<td></td>
</tr>
<tr>
<td><strong>Total Costs</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Translation</td>
<td>170,000</td>
<td>5,200</td>
<td>140</td>
</tr>
<tr>
<td>Post Editing</td>
<td>50,000</td>
<td>50,000</td>
<td>88,400</td>
</tr>
<tr>
<td>Re-Translation</td>
<td></td>
<td></td>
<td>54,400</td>
</tr>
<tr>
<td>Re-Translation Post Editing</td>
<td></td>
<td></td>
<td>16,000</td>
</tr>
<tr>
<td>Proof Reading</td>
<td>30,000</td>
<td>25,000</td>
<td>30,000</td>
</tr>
<tr>
<td><strong>TOTAL COSTS</strong></td>
<td>250,000</td>
<td>80,200</td>
<td>188,940</td>
</tr>
</tbody>
</table>

### Savings Analysis

- **Human vs. MT Value**: 169,800
- **Language Studio™ vs. Google Value**: 108,740
- **Human vs. MT %**: 68%
- **Language Studio™ vs. Google %**: 43%

**Words to Break Even with Language Studio™ = 191,402**
**Total Savings for Full Word Pack Utilization = US$ 1,103,700**

*Rate will vary depending on package purchased. Typical package between US$ 0.005-0.0002 per word.*

*Only content that needed to be 100% retranslated*
How to Measure the Value of MT

**Metrics That Really Count**

- **Productivity** – Words per day per human resource
- **Margin** – 2-3 times the profit margin is commonplace
- **Consistency** – Writing style and terminology
  - MT + Human delivers higher quality than a human only approach
- **Deals**
  - New deals not accessible with a human only approach
  - Deals where you could offer a more competitive bid due to MT than your competitors
  - Deals that would have been lost to a competitor without the advantages that MT offers

**Examples of other “Useful” Quality Indicators**

**Automated Metrics** *(Good indicators, but not absolute)*

- BLEU (Bilingual Evaluation Understudy)
- NIST
- F-Measure (F1 Score or F-Score)
- METEOR (Metric for Evaluation of Translation with Explicit ORdering)

**Manual Quality Metrics** *(Most not designed for MT, more for HT)*

- Edit Distance (Does not take into account complexity of edit)
- SAE-J2450 (Industry specific)

---

**Productivity is the Best Quality Metric**

Raw MT often has a greater number of errors than first pass human translation.

**However:**

Language Studio™ MT is stylised to a specific domain, customer and target audience, so quality is considerably higher than other MT systems.

**This means that:**

1. MT errors are easy to see and easy to fix (i.e. simple grammar).
2. MT provides more accurate and consistent terminology than human translators, especially when more than 1 human works on a project.
3. Human errors may be fewer, but harder to see and harder to fix.

Counting the number of errors only, offers no value as a metric as the complexity of the error is not taken into account.

MT with more errors is often faster to edit and fix than first pass human translations with fewer errors.
Productivity is the Best Quality Metric

MT with more errors is often faster to edit and fix than first pass human translations with “fewer” errors.

- MT errors are easy to see and easy to fix (i.e. simple grammar, capitalization, word order).
- MT provides more accurate and consistent terminology than human translators, especially when multiple translators work on a project.
- Human errors may be fewer in number, but harder to see and harder to fix.
“Good MT” allows more to be translated, faster and at lower cost.
Who Should Manage and Own MT?
The term by its very nature implies that you are required to do something
If you have these skills – contact us – we want to hire you!

It takes Asia Online 12+ months to get a new language engineer up to a high skill level
Why? Because Projects Fail!!

Every MT Vendor has Projects that have Failed

• Most failures are avoidable
• Most fail because
  – The wrong approach to customization was taken
  – The lack of experience of the team doing customization
  – The lack of understanding of how to work with data
  – The lack of skills to be able to diagnose issues
  – The expectation that it was going to be easy
  – Lack of a clear and coherent MT strategy

The “Your MT Department” program is designed to address all these issues and more
Language Studio™ Technology Reduces the Effort to Quality Ratio

Asia Online “Your MT Department” Effort
- Majority of all Technical Tasks
- Custom Engine Development Plan
- Quality Improvement Plan
- Issue Analysis and Resolution
- Terminology Normalization
- Unknown Word Detection and Resolution
- Gap Analysis
- Data Manufacturing
- Rule Creation
- Translation Quality Analysis
- Many more Quality Centric Tasks...

Your Effort
- Term Management
- Engine Definition
- Evaluations

MT Effort

Translation Quality

150-300% Productivity Gain

<20-40% Productivity Gain

DIY Upload TM
Click to Translate
Google
DIY
Your MT Department Enables

• Business Development and Expansion
• Strategic Planning
• More Rapid Exploitation of the Full Commercial Potential of your Custom Engines
• Support and Training for Project Managers
• Monitoring of Technical Performance
• Simplicity, Transparency and Control
• Understanding of Return On Investment and Total Cost of Ownership
Technical Execution and Guidance

- Custom Engine Development
  - Provide guidance and assist project managers
  - Project planning
  - Custom engine training plan
  - Custom engine quality improvement plan
  - Data gathering and acquisition
  - Data manufacturing
  - Data preparation and cleaning
  - Gap analysis
  - Project tracking
  - Pre-translation analysis
  - Post-translation analysis
Language Studio™ provides tools and processes for normalization of terminology

Benefits include cost reductions, faster deliverables, higher customer satisfaction and happier post editors

<table>
<thead>
<tr>
<th>Source</th>
<th>Human Reference</th>
<th>Language Studio™ Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Install a new gasket and the inlet housing with the 3 bolts.</td>
<td>Coloque una junta nueva y el alojamiento de entrada con los 3 pernos.</td>
<td>Instale una junta nueva y el alojamiento de entrada con los 3 pernos.</td>
</tr>
<tr>
<td>Install the brush holder to the generator coil with the 2 screws.</td>
<td>Instale el portaescobillas en la bobina del generador con los 2 tornillos.</td>
<td>Instale el portaescobillas en la bobina del generador con los 2 tornillos.</td>
</tr>
<tr>
<td>Remove the bolt and hole cover from the transaxle.</td>
<td>Quite el perno y la cubierta del orificio de la caja del transeje.</td>
<td>Extraiga el perno y la cubierta del orificio del transeje.</td>
</tr>
<tr>
<td>Remove the clamps and connectors, and then disconnect the wire harness.</td>
<td>Extraiga las abrazaderas y los conectores y, a continuación, desconecte el mazo de cables.</td>
<td>Extraiga las abrazaderas y los conectores, y a continuación desconecte el mazo de cables.</td>
</tr>
</tbody>
</table>
Asia Online takes data security very seriously. We recognize that your translation memories and other linguistic assets are very valuable. Our corporate culture and infrastructure were designed with security and privacy at its core.

Your Data and Your Custom Engines are Private

Data provided by clients will only be used for the purpose of building and improving your own custom engines. We do not share your data with any other clients and is secured by 2048 bit private key encryption. Custom Engines are private to each client. Once a custom engine has been created for a client, the engine can only be accessed by the client or those authorized by the client.

Infrastructure

Asia Online hosts its technical infrastructure in a ISO 27001 certified datacentre and is secured in a private network with world class connectivity and security features.

Contract

Asia Online is very serious about protecting your data and puts its commitments to protect your data into a comprehensive contract. This contract details how your data can and cannot be used by Asia Online and ensures that all intellectual property provided to Asia Online by our clients remains 100% the property of our clients, with no transfer of ownership required.

Policy and Procedures

Asia Online limits access to client data to authorized personnel involved in the customization of a client’s engine. Many of the customization processes are fully automated with no human interaction. All personnel have been interviewed and properly vetted by Asia Online. Each individual employee is subject to a confidentiality agreement and other legal documents in which the employee agrees to protect data.
Getting Started With Machine Translation (MT 101)

Dion Wiggins
Chief Executive Officer
dion.wiggins@asiaonline.net