Fully automatic high quality translation of restricted text

Uwe Muegge
Fully Automatic High Quality Translation of Restricted Text

Uwe Muegge
Global Translation Solutions
Medtronic – Minneapolis/USA
Another great Google tool
Sophisticated machine translation

Machine Translation
Controlled Language Translation Standards

Machine Translation
Controlled Language Translation Standards

Machine translation
Machine translation is an interesting technology for human translators.

It is a fact that machine translation can translate texts very quickly. The question is: Are these machines translations perfect? And the answer is: No. But human translators can correct mistakes in these machine translations very quickly. This web site demonstrates how human translators can use machine translation software in order to improve their translation processes. (link)

Machine translation and controlled language.
Test this efficient translation strategy.
The author of this website has used a controlled language in order to write this website. This website is suitable for machine translation.

Click on the desired language in order to translate this website with the machine translation system from Google:

- German > English
- French > English
- English > French
- English > German
- English > Italian
- English > Portuguese
- English > Spanish
- English > Arabic (Beta)
- English > Chinese (Simplified) (Beta)
- English > Japanese (Beta)
- English > Korean (Beta)
What is crummy machine translation?

Crummy machine translation is MT that’s:
- linguistically not very powerful
- easy to build
- cheap
Challenge

Analysis of previous translations

- Unapproved terminology
- Transposed, missing or otherwise incorrect numbers
- Missing translations
- Missing special characters
- Unapproved formatting
Challenge

Analysis of previous translations

- Lack of quality assurance and quality control steps
- Lack of efficient and effective processes
- Lack of tools optimized for this task
Challenge

Client expectations

Translate product descriptions for use on invoices, customs forms, etc.
- Provide better quality than previous vendor
- Reduce translation lead time
- Cut cost
Analysis of source file

Text characteristics:
- Simple syntactic structure (strings of nouns and adjectives)
- High percentage of word repetition
- Critical importance of integrity of numbers
- Typical for texts in catalogs and inventories
- Applicable to both products and services
- Relevant to almost any organization active in any market
Challenge

Analysis of current tools and processes

Translation Memory Tools

- Segment (sentence) based
  - As each product is unique, fuzzy matches always include irrelevant information
  - Example:
    New string: Lead 9934AB short
    Fuzzy match: Lead 9933AB long

- Fuzzy matches require manual translator intervention
  - Error-prone
  - Time-consuming
  - Expensive
Challenge

Analysis of current tools and processes

Rules-based Machine Translation Tools

- Segment (sentence) based
  - Attempts to produce well-formed sentences, e.g. changes word order, introduces articles, changes inflection
  - Example:
    English string: Catheter 6233L sterile blue
    Spanish string: Azul estéril del catéter 6233L

- Standard machine translation tools require manual translator intervention
  - Error-prone
  - Time-consuming
  - Expensive
Challenge

Analysis of current tools and processes

Ubiquitous type of text that is unsuitable for
- human translation
- computer-assisted translation (TM)
- machine translation (MT)
Solution

Development of translation engine

Specifications

- Find source terms in the source text that are listed in the dictionary
- Replace those source terms in the source text with the target terms listed in the dictionary
- Do not change anything else, e.g.
  - Word order
  - Product numbers
  - Other codes
Solution

Development of dictionary

Dictionary development process

- Leverage existing termbase (7000+ entries)
- Use TERMinator tool to extract new terms
- Have external vendor translate terms
- Import dictionary into translation tool
Solution

Development of dictionary

Dictionary development tactics

- Use canonical forms of entries, i.e. ignore inflections
- Publish list of approved abbreviations to minimize number of variants
Solution

Pricing strategy

Two scenarios

- Pay for every word in a segment minus discounts
  - Translation memory
  - Post-editing machine translation
- Pay only for unique words in a project minus existing terms in dictionary
  - Direct machine translation without post-editing

Example

approx. 250 words
approx. 70 unique words
approx. 15 new terms
Solution

Workflow

1. Extract new terms from source text
2. Translate new terms
3. Import dictionary into MT system
4. Translate source text
Results

Process improvement - financial impact

- Delivers fully automatic high-quality translation
- Requires only a single clerical staff person
- Is scalable to any number of languages
- Reduces translation lead time
- Minimizes translation cost
Summary

Direct MT is alive and well!

- Dramatic improvement in all aspects of the translation process
- Easy-to-implement technology
- Superior translation technology for this text type!