Beyond Google Translate: The future of machine translation

Uwe Muegge
Definition

**machine translation**

type of software application where the program performs the task of creating a target text in another language without human intervention
How can smaller businesses benefit from machine translation?
Strategy
Machine translation for smaller businesses

- Deploy a Small-scale MT Solution
- Manage Terminology
- Control Authoring
Uwe Muegge

Currently
- CSOFT Director, Chief Terminologist
- ISO Member, TC 37
- Monterey Institute of Int'l Studies Prof., Computer-Assisted Translation

Previously
- Medtronic Corporate Terminologist
- J.D. Edwards Localization Specialist
Why should you be using machine translation?
Only a small percentage of what should be translated actually gets translated.
Do you communicate with your clients in the local language - on all channels?
How do you determine if a project is suitable for machine translation?
Leave to humans: Marketing collateral, corporate websites, regulatory filings, etc.
Consider machine translation: Newsletters, knowledge bases, training materials, etc.
What are the strengths and weaknesses of Google Translate?
Definition

**statistical machine translation**

Type of machine translation that uses statistical models, which are based on the analysis of large bilingual corpora, to generate sentences in the target language.
Statistical Machine Translation

- **Bilingual training corpus**
- **Target language corpus**

**Translation Model**
- Calculates the probability of a given source-target sentence pair

**Language Model**
- Calculates the probability of a given sentence occurring in the target language
Google Translate
What's to like

- Free to use
- Very fast
- Fairly accurate
- Vendor independent
- Free API
- Large user base
Google Translate
What's not to like

- No control over terminology
- Privacy concerns
- No control over style
Google Translate allows you to specify only two things:
1) text
2) language direction
If you want more control over your MT system, what are the options?
Google Translate Alternatives

- **User friendly, inexpensive**: Deploy a rule-based (e.g. Systran) MT system
- **Tool expensive**: Deploy a commercial (e.g. Language Weaver) statistical MT system
- **Staff expensive**: Deploy a free (e.g. MOSES) statistical MT system
Definition

rule-based machine translation

type of machine translation that uses syntactical rules for analyzing sentences in the source language and for generating sentences in the target language
Rule-based Machine Translation

- **Analysis Module**: Generates an abstract structure of the source-language sentence.
- **Transfer Module**: Transfers source-language structure into target-language structure.
- **Generation Module**: Generates target language sentence from abstract structure.
Rule-based Machine Translation

Analysis tree shows syntactical structure of source sentence

Transfer tree shows syntactical structure of target sentence
Rule-based MT
What's to like

- Very low cost
- Can be operated by non-linguists
- Fairly accurate
- Customizable
- Works right out of the box
- Some systems feature a translation memory
How can you customize a rule-based machine translation system?
Creating/harvesting project-specific terminology
Definition

terminology management
process of controlling key vocabulary
Benefits
Performing Terminology Management

- Meeting a Critical User Requirement
- Compliance with Industry Best Practices
- Simplified Translation Editing and Reviewing
- Simplified Terminology QA after Translation
- High-quality Machine Translation
- Automatic Terminology Lookup
- Consistency across Functional Areas
- Consistency across Different Versions
- Protection of Intellectual Property
- Simplified Source Writing and Reviewing
Free, web-based solutions like TermWiki make terminology management easier than ever.
Terminology Component of a Rule-based MT System

Adding terminology is as easy as importing a bilingual table.
What else can you do to optimize translation quality?
Control authoring
Sample Writing Guidelines

- Write sentences that are shorter than 25 words
- Write sentences in the active form
- Write sentences that repeat the noun instead of using a pronoun
- Write sentences that use articles to identify nouns
Quick Demo
Der Autor dieser Website benutzte eine kontrollierte Sprache zum Verfassen dieses Textes in der englischen Sprache. In einem zweiten Schritt übersetzte ein maschinelles Übersetzungssystem den englischen Text in die deutsche Sprache.

The author of this website used a controlled language for writing this text in the English language. In a second step translated a machine translation system this English text into German Language.
Strategy
Machine translation for smaller businesses

Deploy a Small-scale MT Solution
Manage Terminology
Control Authoring
What are the steps to implementing a rule-based MT program?
Machine Translation
Rule-based pilot

1. Establish a translation productivity baseline
2. Select a suitable source text, determine language pair
3. Deploy an MT system, ideally with translation memory
4. Extract terminology from the source text
5. Have terminology translated and reviewed
6. Import terminology/customize MT system
7. Machine-translate your document(s)
8. Post-edit the machine-generated translations
9. Solicit feedback from readers and post-editors
10. Measure productivity gains, optimize MT system
You can make machine translation work for you!
Thank you for your attention!

Uwe Muegge
uwe.muegge@csoftintl.com
Twitter: @UweMuegge

CSOFT™
One World, One Market