[*Translating and the Computer 28*, November 2006 (London: Aslib, 2006)]

Fully Automatic High Quality Machine Translation of Restricted Text: A Case Study

Uwe Muegge
(uwe.muegge@medtronic.com, uwe.muegge@miis.edu)

Global Translation Solutions
Medtronic, Inc.
Minneapolis, MN - USA

Graduate School of Translation and Interpretation
Monterey Institute of International Studies
Monterey, CA - USA

Abstract

Medtronic is currently in the process of consolidating multiple distributed legacy product databases into one centrally managed SAP database. With a nine-figure budget, the Centerpiece effort is the largest and most visible IT project Medtronic has ever undertaken. One crucial part of this project is the translation - into eight languages - of existing descriptions for 50000 products, as well as approx. 200 new descriptions that are being added to the database every week.

With both normalized source text and comprehensive, authoritative terminology in place, Medtronic is in an excellent position to use machine translation to produce translations of these product descriptions at the push of a button. This presentation illustrates the processes that allow Medtronic to produce translations in-house, instantly, at higher quality than previous human translations, and at a fraction of the cost of human translation.

1. Project Background

1.1. Product Database Consolidation

With offices in more than 120 countries and annual revenues of more than 11 billion dollars, Medtronic is the world leader in medical technology. Since 2004, Medtronic has been consolidating multiple distributed legacy product databases into one centrally managed SAP database. With a budget in excess of 200 million dollars, the Centerpiece effort is the largest and most visible IT project Medtronic has ever undertaken. One critical part of this project is the translation into multiple languages of the existing descriptions for 50,000 products, plus approximately 200 new descriptions that are added to the product database every week.

2. Existing Translation Environment

2.1. Solution for Technical Literature

The workflow for technical literature is highly automated and integrates a proprietary content management system (MAPS) with a customized, off-the-shelf workflow solution (Trados/SDL TeamWorks). This solution is geared towards translating structured XML output that is characterized by the following typical translation memory match rates:

- 70% perfect matches
- 23% fuzzy matches
- 7% no matches

This type of translation project is typically a low-volume job that is handled by in-house translators.

2.2. Solution for Marketing Literature

The workflow for marketing literature is quality driven and uses a number of labor intensive processes to ensure the highest degree of customer satisfaction possible. Marketing translation documents are translated in industry-standard translation memory systems. This solution is characterized by the following typical translation memory match rates:

- 10% perfect matches
- 20% fuzzy matches
- 70 % no matches

This type of translation project is typically a medium-volume job that is handled by a small group of external translators, each of whom receives product training prior to each translation project.

2.3. Product Database Translation Does Not Fit Existing Workflows

Translating Medtronic's large product database is not a good fit for the two existing workflows. The automated workflow for technical literature relies heavily on leveraging matches from existing translation memories. As there would have been no translation memory matches for the initial round of translations, our human technical translators would have been overwhelmed by the sheer volume of work. Our quality-driven workflow for marketing literature was also not usable as marketing translators have a skill set that differs greatly from the one required to perform the monotonous task of translating tens of thousands of product descriptions.

3. Designing the Product Description Translation Environment

3.1. Client Sets Aggressive Goals

In the discussions with the owners of the product database, the management team defined the following goals for the new translation process:

- Improve translation quality over previous translation efforts
- Reduce the cost of translation
- Reduce the turnaround time
- Reduce the involvement of marketing staff in the translation process
- Automate the translation process as much as possible

3.2. Issues with Conventional Translation Tools

There was little question that our primary translation tool, the Trados Workbench translation memory, would not allow us to make the breakthrough improvement in the translation process we needed to succeed. However, we were surprised to find that even with our newest tool, the Systran machine translation system, we could not meet the client requirements for this project.

Issues with our Trados translation memory environment include:

- Requires human interaction for every segment that is not a 100% match
- Introduces errors because fuzzy matches include wrong product number
- Does not force translators to use approved terminology

Issues with our Systran machine translation environment include:

- Introduces errors because it misinterprets incomplete sentences
- Requires human interaction for post-editing

3.3. Direct Machine Translation Is the Solution of Choice

Direct machine translation is the least sophisticated approach to automated language processing. Unlike rules-based systems like Systran, direct machine translation systems do not perform a grammatical analysis of the source sentence. Also, direct machine translation systems have no concept of inflection. In fact, all these systems do is perform a word-for-word substitution, which is exactly what we were looking for.

Benefits of direct machine translation technology include:

- Generates translations of completely new text
- Translates only words that are in the dictionary
- Translates large volumes of text almost instantly
- Processes many languages in the same system
- Is very inexpensive to purchase and maintain

4. Implementing the Product Description Translation Environment

4.1. Controlling Input

As the consolidation of legacy databases progressed, it became obvious that there was very little common ground in the linguistic features of the product descriptions in these databases. Some entries consisted entirely of numerical dimensions, where others contained no numerical information at all. Abbreviations and acronyms posed another serious challenge, e.g. '`ster`' stood for 'sterilized' in one database and 'steroid' in another.

We took the following steps to control the source text:

- Implement a standard for writing product descriptions
- Implement a list of approved abbreviations and acronyms
- Implement an automatic checking tool
- Implement a process where all non-compliant product descriptions are returned to the business units for correction.

4.2. Managing Terminology

One of the strengths of the direct machine translation approach is its consistent translation of terminology. The fact that direct machine translation systems only translate words that are in their dictionary mandates a comprehensive approach to terminology management.

The stages involved in our terminology management effort include:

- Use TERMinator automatic terminology extraction tool
- Have external vendor translate English terms
- Have Medtronic in-country subject matter experts review translated terminology
- Import reviewed terminology into machine translation dictionary

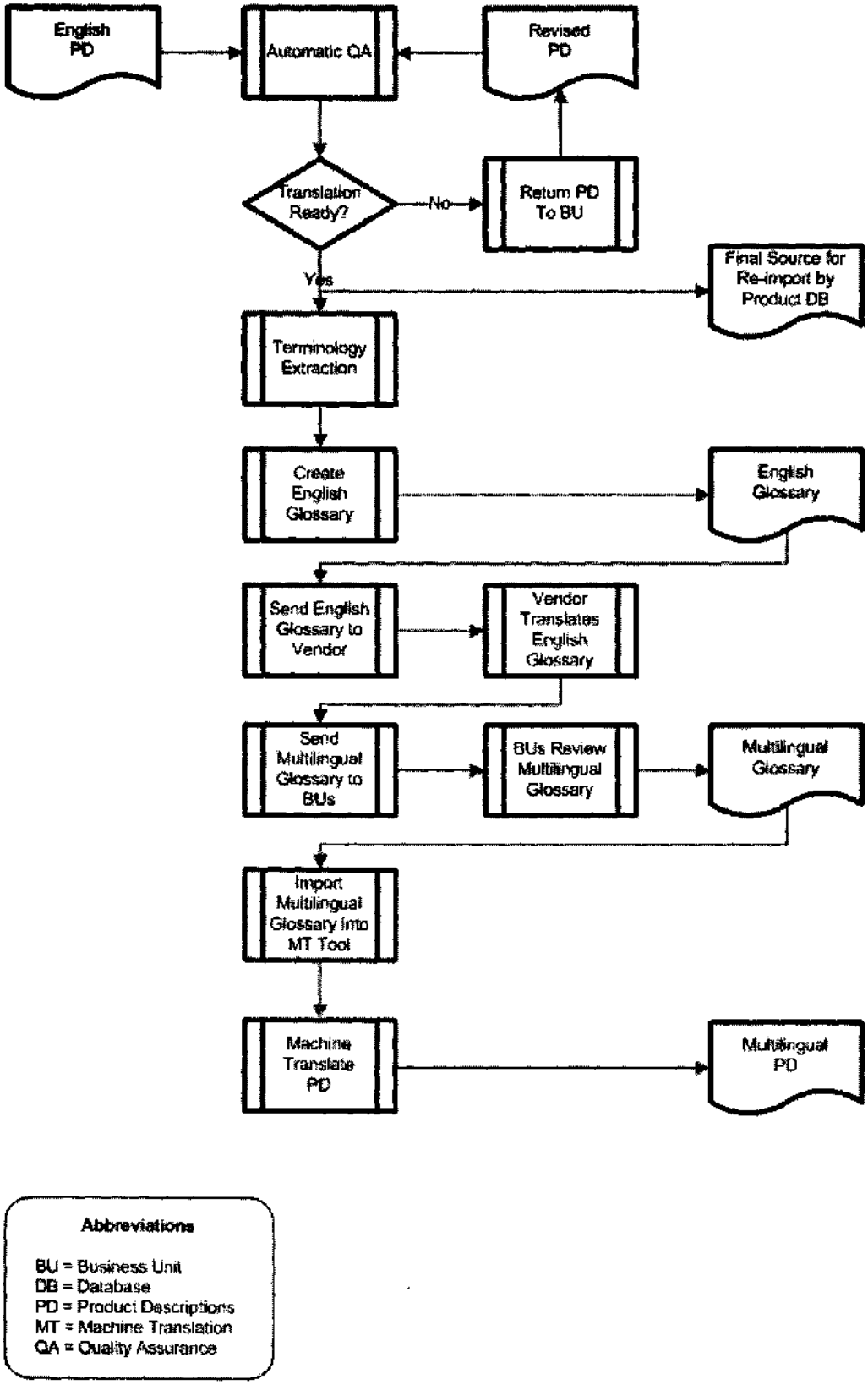

Figure 1: *Workflow for machine-translating product descriptions*

4.3. Benefits of this Translation Solution

After using these tools and processes for almost two years in the translation of product descriptions, the Global Translation Services group within Medtronic has met if not exceeded the expectations of our various internal clients. Our process for translating product descriptions consistently delivers a translation product that has the following characteristics:

- Extremely short turnaround time (in some cases, translations were returned within 15 minutes of receipt of order)
- Highly accurate in the use of preferred client terminology
- All translation work can be performed internally
- Cost of translation is one order of magnitude lower than traditional human translation

5. Selected Publications

Muegge, U. (2006). "Rules for machine translation." http://www.muegge.cc/machinetranslationcontrolledlanguages/machine-translation-controlled-languages.htm.

Muegge, U. (2005). Translation contract: a standards-based model solution. Bloomington, AuthorHouse.

Muegge, U. (2002). "Lokalisierung und Maschinelle Übersetzungssysteme." Lokalisierung von Technischer Dokumentation. M. Tjarks-Sobhani and J. Hennig. Stuttgart, Gesellschaft für technische Kommunikation (tekom). 6: 101-121.

Muegge, U. (2002). "Möglichkeiten für das Realisieren einer einfachen Kontrollierten Sprache." Lebende Sprachen 47(3): 110-114.

Muegge, U. (2001). "The best of two worlds. Integrating machine translation into standard translation memories. A universal approach based on the TMX standard." Language International 13(6): 26-29.