Introduction to Terminology Management for Localization

Sorrel Ritter

February 27th, 2014
use mouse to open window
Text without Terminology

The is a by which from a is absorbed by, and is re-radiated in all. Since part of this is back towards the and the lower of the what it would be in the at the of largely passes through the to warm the which then emits this at the lower of.
Text with Terminology

The greenhouse effect is a process by which thermal radiation from a planetary surface is absorbed by atmospheric greenhouse gases, and is re-radiated in all directions. Since part of this re-radiation is back towards the surface and the lower atmosphere, it results in an elevation of the average surface temperature above what it would be in the absence of the gases. Solar radiation at the frequencies of visible light largely passes through the atmosphere to warm the planetary surface, which then emits this energy at the lower frequencies of infrared thermal radiation.
Result

- greenhouse effect
- process
- thermal radiation
- planetary surface
- atmospheric greenhouse gases
- directions
- re-radiation
- atmosphere
- elevation
- average surface temperature
- etc.
Terminology (ISO 1087-1/2000:10)

• Set of designations belonging to one special language
Terminology work/management

- ISO 1087-1/2000:10
  Work concerned with the systematic collection, description, processing and presentation of concepts and their designations

- Terminology work consists of establishing specialized vocabularies, that is, tying designations to their respective concepts in a particular subject field
Why is terminology management so important…
Synonyms or different concepts?

- Within one text or across different texts/media

**English:**

- Hill hold
- Hill-holder
- Hill Hold Control (HHC)
- Starting-off aid (Hill Hold Control)
- Hill hold assist function
- Hill-start assist
## Synonyms or different concepts?

<table>
<thead>
<tr>
<th>Term [EN]</th>
<th>Meaning</th>
<th>Part of speech</th>
<th>Context</th>
<th>Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strike</td>
<td>To hit</td>
<td>Verb</td>
<td>To inflict a blow</td>
<td>打つ</td>
</tr>
<tr>
<td>Strike</td>
<td>To engage in a work stoppage</td>
<td>Verb</td>
<td>A work stoppage caused by the mass refusal of employees to work.</td>
<td>ストライキする（同盟罷業）</td>
</tr>
<tr>
<td>Strike</td>
<td>A pitch that ‘passes over home plate…..’</td>
<td>Noun</td>
<td>A baseball term</td>
<td>ストライク</td>
</tr>
<tr>
<td>Strike</td>
<td>To find something of value…</td>
<td>Noun</td>
<td>e.g. a Gold Strike</td>
<td>大当たり</td>
</tr>
</tbody>
</table>
Abbreviations

TMS

• Translation Memory System
• Translation Management System
• Terminology Management System

• The Muppet Show…
• Too Much Stuff…
Orthographic variants

Reference number
• Ref. No.
• Reference No.
• Ref #
• Ref.-No.
• Ref No.
• Ref. No.: 
• Ref. Number
• Ref. Number:
International branding

Nissan "Pivo"
Czech: pivo = beer

Mitsubishi "Pajero"
Spanish: pajero = wet blanket

Toyota "MR2"
French: [mɛʁd] merde

Nissan "Juke"
Hebrew: juke = cockroach

Source: Diana Brändle
© 2014 Net-Translators. All rights reserved.
Cultural issues

USA

CHINA

Example "1969" taken from MultiLingual April/May 2012 "International branding errors cause trouble"
Terminology management pain curve

Source: Multilingual April/May 2007; Terminology: ignore it at your peril – Keiran J. Dunne
Collaboration

• Team
  – Marketing
  – Development
  – Technical communications

• Accessibility
  – Editors
  – Engineers
  – Terminologists
  – Producers
  – Language Service Provider (LSP)
  – Linguists
IMPORTANT

• Who has the last word on terminology within the company… otherwise discussions may be endless, and cause terminology projects to fail.
Ongoing process

• Adding new terms
• Adding new languages
• Adding additional information e.g. pictures
• Changing translations/incorporating feedback from users
• Changing the status of a term e.g. allowed, forbidden, pending…
• Separating or combining terminology resources
• Converting terminology resources for use within other tools
Semantic triangle

- **Thought/Concept**
  (unit of thought, unit of knowledge)

- **Designation/Term**
  (sign, term, symbol, formula)

- **Object/Referent**
  (material, immaterial or imagined objects)

"knife"
Object (ISO 1087-1/2000:2)

• Anything perceivable or conceivable
• Note: Objects may be:
  – Material e.g. pencil, albatross, forest
  – Immaterial e.g. a dream, pH
  – Imagined e.g. fairy
Concept (ISO 1087-1/2000:2)

- Unit of knowledge created by a unique combination of characteristics
  - Individual concept
  - General concept
- Note: Concepts are not necessarily bound to particular languages. They are, however, influenced by the social or cultural background which often leads to different categorizations.

- Representation of a concept by a sign which denotes it
- The designation acts as a synthesis of the definition
- Note: In terminology work three types of designations are distinguished:
  - symbols
  - appellations
  - terms
The ideal situation

• Terminological consistency:
  – Only one designation per concept, no synonyms
  – Only one concept per designation, no ambiguities
Imbalances between designations and concepts

- Synonymy – one concept, but several designations
- Ambiguity – one designation, but several concepts

Handbag
Pocketbook
Bat
What is a term

A term, or terminology unit, is the name or designation of a concept in a particular subject field. A term may be:

- a word (a general language word used in a specialized way or a new word created to represent a concept in a specialized field)
- an expression
- a symbol
- a chemical or mathematical formula
- a scientific name in Latin
- an acronym
- an initialism
- the official title of a position, organization or administrative unit
What terms should be managed

Following are a few considerations to help focus terminology efforts where they matter most:

- Safety of customer
- Survival of company
- Survival of product
- Company specific
- Industry specific
- Efficiency
What terms should be managed (cont.)

Following are criteria for identifying terms for inclusion in a terminology database (TermBase):

- Terminologization
- Confusability
- Specialization
- Frequency and distribution
- Novelty
- Visibility
- System
- Standardization

Source: Barbara Inge Karsch
What terms should be managed (cont.)

- Each one of the criteria weighs differently for each term candidate, and most of the time several criteria apply.
- A terminologist, content publisher or translator has to weight these criteria and make a decision quickly.
- No two people will come up with the same terminology list.
How many terms should be included

A few key criteria that help determine the number of terms to document in a terminology database:

- **Size of the project**: Small, medium, large, extra-large
- **Timeline**: Six days or six months to the deadline
- **Version**: Version 1 or 2 of the product, or 7 or 8
- **Number of terms existing in the database already**: Is this the first time that terminology has been documented
- **Headcount**: Number of people that will be documenting terms and how much time they can devote
- **Level of complexity**: New products or features, subject matter content higher than normal, small update

Source: Barbara Inge Karsch
Term extraction

- Manual term extraction
- Tool-based term extraction
- Three approaches in tools:
  - Statistical: strength - language-independent; weakness - "noise" and "silence"
  - Linguistic: weakness – language-dependent
  - Combined linguistic and statistical approach
How term appears in TermBase

Terms should be entered in *canonical form* (often called *normal form* or *standard form*). The canonical form of an object is a standard way of presenting that object:

- Normally in lowercase (but: Elizabeth, Apple)
- Nouns normally in singular (but: trousers, reading glasses)
- Verbs normally in infinite form, not with infinite particles (to)
- Multiword terms in spoken/written order
- Nouns not with articles (the, an)
- Spelled correctly and consistently
Data categories

- Subject field/domain
- Product name
- Department name
- Project name
- Definition
- D_source
- Language
- Geographical usage (locales e.g. UK, US)
- Part of speech
- Usage status (Preferred, Forbidden, Proposed…)
- Gender
- Context
- Context_source
- Illustration
- Illustration_source
TermBase

Entry = concept

Language

Term

Source: Prof. Klaus-Dirk Schmitz
Tools

• Controlled Language Checker (CLC)/Terminology checking (Authoring Management System (AMS))
  – Acrolinx
  – crossAuthor (across)
  – MindReader (STAR)
  – SDL Authoring Assistant
Tools (cont.)

• Term extraction, many commercial and free tools:
  
  **Commercial**
  – Acrolinx
  – Across
  – LogiTerm
  – SDL MultiTerm Extract
  – SynchroTerm

  **Free**
  – FiveFilters
  – Similis
  – TerMine
Tools (cont.)

• Terminology Management Systems (TMS)
  – crossTerm
  – i-Term
  – MemoQ
  – qTerm
  – SDL MultiTerm
  – TermWeb
  – TermWiki (Pro)

• Terminology workflow tools
  – Integrated features or add-on tools
Preparing a TermBase

• Creating source terminology list
• Client approval
• Translation of terminology list
• Client approval
• Generation of MultiTerm (MT)
MT in CAT tool

<table>
<thead>
<tr>
<th>Week</th>
<th>Woche</th>
</tr>
</thead>
<tbody>
<tr>
<td>100%</td>
<td></td>
</tr>
</tbody>
</table>

- EN to DE-DE - Translation Results

- Term Recognition

- Week:
  - Field: week
  - Description: "Week"
  - Part of Speech: "Noun"
  - Plural Name: "Wochen"
Equivalence (ISO 1087-1/2000:9)

• Cookie

South African English

US English
Equivalence (ISO 1087-1/2000:9) (cont.)

- Robot

South African English

US English
Conclusions

• Proper terminology is essential for successful translation
• It is easy to start
• Resources are available to help you, for example:
  TermNet – International Network for Terminology
  http://www.termnet.org/
  International Terminology Summer School
• Even if you are currently translating you can still rectify the situation
Thank you for your attention!