About Heather Hedden

- Taxonomy consultant with Project Performance Corporation
  www.ppc.com
- Continuing education instructor with Simmons College Graduate School of Library and Information Science
  alanis.simmons.edu/ceweb
- Author of *The Accidental Taxonomist* (Information Today, Inc., 2010)
  www.accidental-taxonomist.com

Broad experience creating taxonomies for:
- corporate web sites and intranets
- consumer web sites
- document management in SharePoint
- integration within an enterprise search software product
- periodical index databases
Introduction to Taxonomies

A Taxonomy is…

• An authoritative, restricted list of terms (words or phrases)
• Where each term is for a single unambiguous concept
• Governed by policies for who, when, and how new terms can be added
• Arranged in some structure or groupings
• To support indexing/tagging/categorizing/metadata management of content to facilitate retrieval/findability

May also include..

• Synonyms as cross-references to terms
• Additional relationships between terms (broader, narrower, related)
• Notes or additional data for each term
Hierarchical taxonomy examples

Broader terms & Narrower terms in “trees”
Faceted taxonomy examples

The user selects one value from each of multiple facets/attributes/filters and searches on them in combination.
Ecommerce taxonomies
(hierarchical + faceted)
Introduction to Taxonomies

Taxonomy Benefits

1. Structured as hierarchies or facets…
   Organizes information into a logical structure
   - Helps people browse or navigate for information
     “Guided search”

2. Managed a controlled vocabulary…
   Brings together different wordings (synonyms) for the same concept and disambiguates terms
   - Helps people search for information by different names
   - Helps people retrieve matching concepts, not just words
   Picks up/enhances where search alone fails
Introduction to Taxonomies

Taxonomy Benefits (as a controlled vocabulary)

- Helps users search for information by different names
  - A controlled vocabulary gathers synonyms, acronyms, variant spellings, etc.
  - Without a controlled vocabulary keyword searches would miss some relevant documents, due to use of different words (e.g. *Automobiles*, instead of *Cars*)

- Helps users retrieve matching concepts, not just words
  - A single term may have multiple meanings.
  - Excludes document with mere text-string matches (e.g. *monitors* for computers, not the verb “observes”)
Introduction to Taxonomies

Taxonomy Term Relationships

1. Equivalence
A “preferred term” and its various “nonpreferred term” variants (“synonyms”), pointing to it as cross-references to use (USE) and is used from (UF)
(e.g. Cars USE Automobiles)

2. Hierarchical
A broader term (BT) and its more specific term, instance, or component part narrower term (NT)
(e.g. Motor Vehicles NT Automobiles)

3. Associative
Any other nonhierarchical, related term (RT) relationship
(e.g. Automobiles RT Automobile Repair Services)
1. Multilingual Taxonomy Goals
2. Multilingual Taxonomy Design
3. Taxonomy Translating
Bilingual/Multilingual Taxonomies can enable:

1. A user to search and retrieve content that is in multiple languages through a single taxonomy in their own language.
Bilingual/Multilingual Taxonomies can enable:

2. Different users who speak different languages to search the same body of content (in one other language), each using a taxonomy in the user interface in their native language.
Bilingual/Multilingual Taxonomies can enable:
3. Different users who speak different languages to search the same body of content that is in multiple languages.
Goals #1 or #2: Users of one language can access content in a different language.
- Taxonomy in one language with equivalent translated terms
- The taxonomy needs to function in only one direction.

Goal #3: Multilingual users can access multilingual content.
- Fully multilingual taxonomy or distinct taxonomies for each language linked at equivalent-meaning terms
- The taxonomy needs to function in both/all language directions.
Multilingual Taxonomy Goals

Different scenario: Multiple language taxonomies, each connected to its own language content, such as for separate web sites.

Spanish speaker

French speaker

German speaker

Multiple, different language UIs.
Outline

1. Multilingual Taxonomy Goals
2. Multilingual Taxonomy Design
3. Taxonomy Translating
Design the multilingual taxonomy to meet the taxonomy goals.

In a one-direction translated taxonomy:
• The language of the searcher has structure to display.
• The language of the content may not need structure.
• Translations may be in one direction (user/display term *may be used for* content/index term, not vice versa).

For a fully bidirectional multilingual taxonomy:
• Both language taxonomies need structure.
• Translations must be exact matches in both directions.

For separate taxonomies in different languages:
• Taxonomies are not translated but each created and managed separately.
Multilingual Taxonomy Design

Dedicated taxonomy/thesaurus management software tools provide varying multilingual capabilities:

1. Customized text field used for term translations
   - No vocabulary control of second language(s)
2. Translations as a controlled vocabulary flat list, linked at each term
   - Vocabulary control of second language(s)
   - Taxonomy structure of primary language inferred
3. Second language taxonomy mirroring first, linked at each translated term
   - Vocabulary control of second language(s)
   - Copying taxonomy structure of primary language
4. Multiple taxonomies in different languages linked at equivalent term translations
   - Each language may have its own structure (requires additional work to build)
1. Customized field used for term translations
2. Translations as a controlled vocabulary flat list, linked at each term
3. Second language taxonomy mirroring first, linked at each translated term. Inter-term relationships replicate.
4. Multiple taxonomies in different languages linked at equivalent term translations. Inter-term relationships may differ.
Multilingual Taxonomy Design

Dedicated taxonomy/thesaurus management software tool screenshot examples from:

• Data Harmony Thesaurus Master (Access Innovations, Inc.)
• Synaptica (Synaptica, LLC)
• MultiTes (Multisystems)

Other tools also provide similar capabilities.
Method #1: Create user-defined text field and enter translation

Data Harmony Thesaurus Master
Multilingual Taxonomy Design

Method #1
Synaptica
Method #2 or 4: Link equivalent terms in different language by user-defined “alternative language” reciprocal relationship

Data Harmony
Thesaurus Master
Method #2 or 4: Link equivalent terms in different language by user-defined associative relationship.

Synaptica
Multilingual Taxonomy Design

Method #2 or 4 alternate: Create separate taxonomies, and link equivalent terms in different language by same Concept ID.
Method #3: Create second language taxonomy mirroring first, linked at each translated term. Inter-term relationships replicate.

**MultiTes**
Multilingual Taxonomy Design

- Translations of a term display as another kind of relationship.
- Similar to equivalence, but both languages are preferred and none is nonpreferred

From the bilingual European Training Thesaurus http://libserver.cedefop.europa.eu/ett
Outline

1. Multilingual Taxonomy Goals
2. Multilingual Taxonomy Design
3. Taxonomy Translating
Taxonomy Translating

- Taxonomy translations are typically created from scratch, translating each term.
- It is also possible to map and existing/separately created foreign language taxonomies to another, if their coverage is nearly identical.

- For Goals #1 or #2 (*Users of one language accessing content in a different language*) translations may suffice
- For Goal #3 (*Multilingual users accessing multilingual content*) mapping separately created taxonomies in each language may be better.
Taxonomy Translating

• Matches are for concepts, not terms.
  – Translations are for the concept and not necessarily for the preferred term.

• Nonpreferred terms may vary.
  – Some can be translated
  – Some cannot be translated
  – Additional nonpreferred terms may be created in the second language(s)
Translating taxonomies/thesauri is different from translating documents.

- Pay by hour/project, not by word.
- Translators should have experience with translating in both directions.
- Translators should be familiar with using taxonomies, if not also taxonomists.
- If not using a translator who is also a taxonomist, have a taxonomist/information-specialist native speaker of target languages review the translated taxonomy.
Taxonomy Translating

Taxonomy Translation/Localization Issues

• Lack of an equivalent translation
• A term in one language having two meanings with two terms in another language (seguridad = safety & security)
• Term length
• Use of definite articles
• Use of abbreviations
• Use of plural
• Use of capitalization
• Alphabetizing sorting rules
Translation projects end, but taxonomy management does not. Taxonomy management issues:

- Taxonomy growth
- Taxonomy change
- Taxonomy management/ownership responsibility
- Merging or combining additional taxonomies

Translations/additional language versions will need frequent reviewing and updating.
Conclusions

• Multilingual taxonomies can serve various goals.
• Design the taxonomy with the goal(s) in mind.
• There are advantages/disadvantages to translating a taxonomy vs. mapping pre-existing taxonomies of different languages.
Heather Hedden  
Senior Analyst  
Project Performance Corporation  
Carlisle, MA  
Corporate office:  
1760 Old Meadow Rd., McLean, VA 22102  
Heather.Hedden@ppc.com  
www.ppc.com  
www.accidental-taxonomist.com  
(703) 462-3746