Creating Multilingual Taxonomies

Heather Hedden
Senior Analyst

Gilbane Conference, Boston, MA, December 1, 2011

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

USA TODAY Content Tree

- USA TODAY Content Tree
  - ⊞ news 218,440 stories
  - ⊞ travel 38,426 stories
  - ⊟ money 57,615 stories
    - _ 2 stories
    - ⊞ advertising 799 stories
    - ⊞ autos 3,325 stories
    - biztravel 734 stories
    - ⊞ books 463 stories
    - ⊞ companies 3,344 stories
    - ⊟ economy 16,598 stories
      - confidence 49 stories
      - construction 7 stories
      - employment 336 stories
      - ⊟ fed 357 stories
        - beigebook 19 stories
        - rates 85 stories
      - gdp 37 stories
      - ⊞ housing 1,739 stories
      - income 49 stories
      - inflation 92 stories
      - production 80 stories
      - productivity 22 stories
      - services 23 stories
      - survey 20 stories
      - trade 157 stories
    - gallery 3 stories
    - ⊞ industries 10,120 stories
    - ⊞ jobcenter 189 stories
    - ⊞ markets 2,844 stories

Site Content and Structure

Back to 'Human Resources'

Site Content and Structure

Refresh all

- Breeze
  - ⊞ Breeze Search Center
  - ⊞ communications
  - ⊟ Departments
    - ⊞ Accounting
    - ⊞ Commercial Asset Management
    - ⊞ Construction
    - ⊞ Corporate Communications
    - ⊞ Development
    - ⊞ Engineering
    - ⊞ Environmental
    - ⊞ Executive Management
    - ⊞ External Affairs
    - ⊞ Facilities
    - ⊞ Finance
    - ⊟ Human Resources
      - ⊞ Announcements
      - ⊞ Benefit Links
      - ⊞ Benefits
      - ⊞ Calendar
      - ⊞ Documents
      - Employee Forms
      - Employee Performance Appraisal
      - Employee Self Assessment
      - ⊞ Handbook & Policies
      - ⊞ Images
      - ⊟ New Hire
        - ⊞ Visa Corporate Card
        - ⊞ Administrative Forms
        - ⊞ Benefits
        - ⊞ Payroll
        - ⊞ I-9
      - ⊞ Pages
      - ⊞ Payroll Forms
      - ⊞ Performance Management Process
      - ⊞ Photo Library
      - ⊞ Shared Documents
      - ⊞ Workflow Tasks
    - ⊟ Information Technology

Human Resources

Up New ▾

Type Title

Announ...
Benefi...
Benefi...
Calend...
Docum...
Emplo...
Emplo...
Emplo...
Handb...
Image...
New H...
Pages...
Payrol...
Perfor...
Photo...
Share...
Workf...
defaul...

Hierarchical taxonomy examples

Broader terms & Narrower terms in "trees"

Course
- Main Dishes (15504)
- Desserts (7530)
- Side Dishes/Vegetables (6182)
- ⊞ Show More

Convenience
- Entertaining (23804)
- Make-Ahead (13917)
- Quick/Easy (13186)
- ⊞ Show More

Cost Per Serving
- $1 and Under (388)
- $1.01 to $2 (394)
- $2.01 to $3 (250)
- $3.01 to $4 (94)
- $4.01 and Up (28)

Cuisine
- American (28614)
- Italian (3129)
- New American (2370)
- ⊞ Show More

Main Ingredient
- Vegetables (11246)
- Fruits (6297)
- Poultry (5287)
- ⊞ Show More

Dietary Consideration
- Meatless (11299)
- Low Cholesterol (7534)
- Low Saturated Fat (7444)
- ⊞ Show More

Cooking Method
- Bake (12470)

NARROW YOUR RESULTS

Carrier
Verizon Wireless (50)
Sprint (50)
AT&T (60)
T-Mobile (32)

Customer Reviews
Top-Rated (26)

Price
Free (75)
$0.01 - $49.99 (33)
$50 - $99.99 (31)
$100 - $149.99 (9)
$150 - $199.99 (15)
$200 - $249.99 (9)
$250 - $499.99 (12)
$500 - $749.99 (6)
$750 - $999.99 (2)

Device Brand
Samsung (44)
Apple (33)
HTC (29)
LG (28)
BlackBerry (22)
Pantech (12)
Motorola (11)
Kyocera (4)
See all...

Operating System
Android (57)
BlackBerry (20)
iPhone (33)

Narrow Your Search

+ **Search Within Results**

− **Locations Served**
Arizona
Arkansas
California - North
California - South
Colorado
[+] More

+ **Search Within # Miles**

− **Company Type**
Manufacturers
Custom Manufacturers
Distributors
Service Companies
Manufacturers' Reps

+ **Certifications**

− **Ownership**
Minority-Owned
Woman-Owned
Veteran-Owned

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)

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

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**)
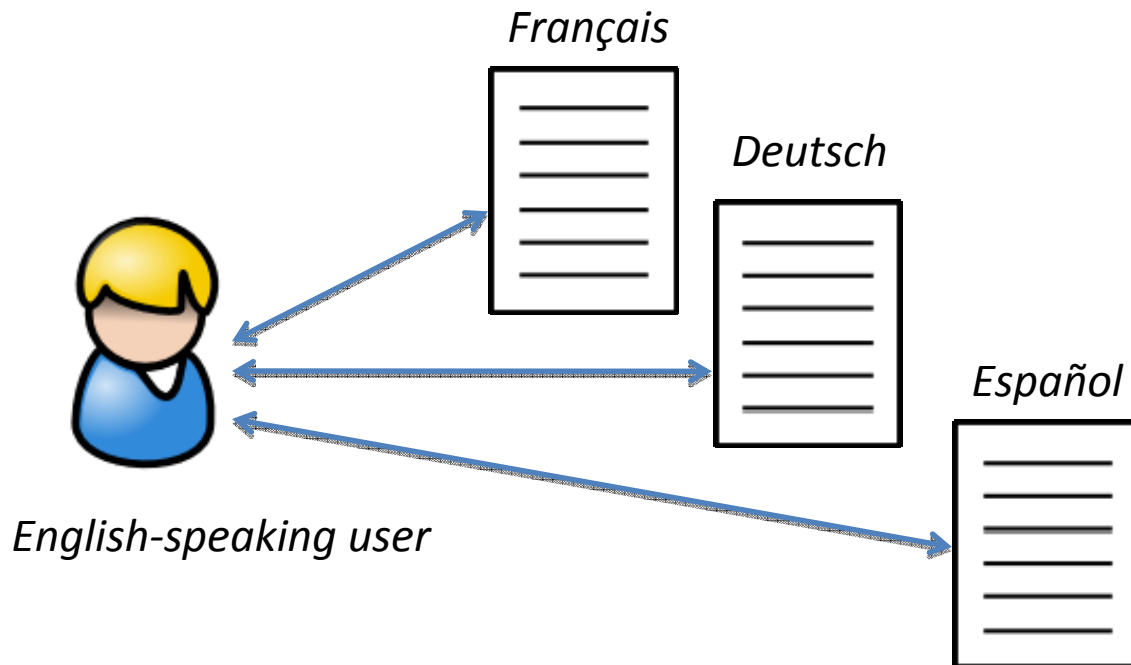
Outline

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

Multilingual Taxonomy Goals

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
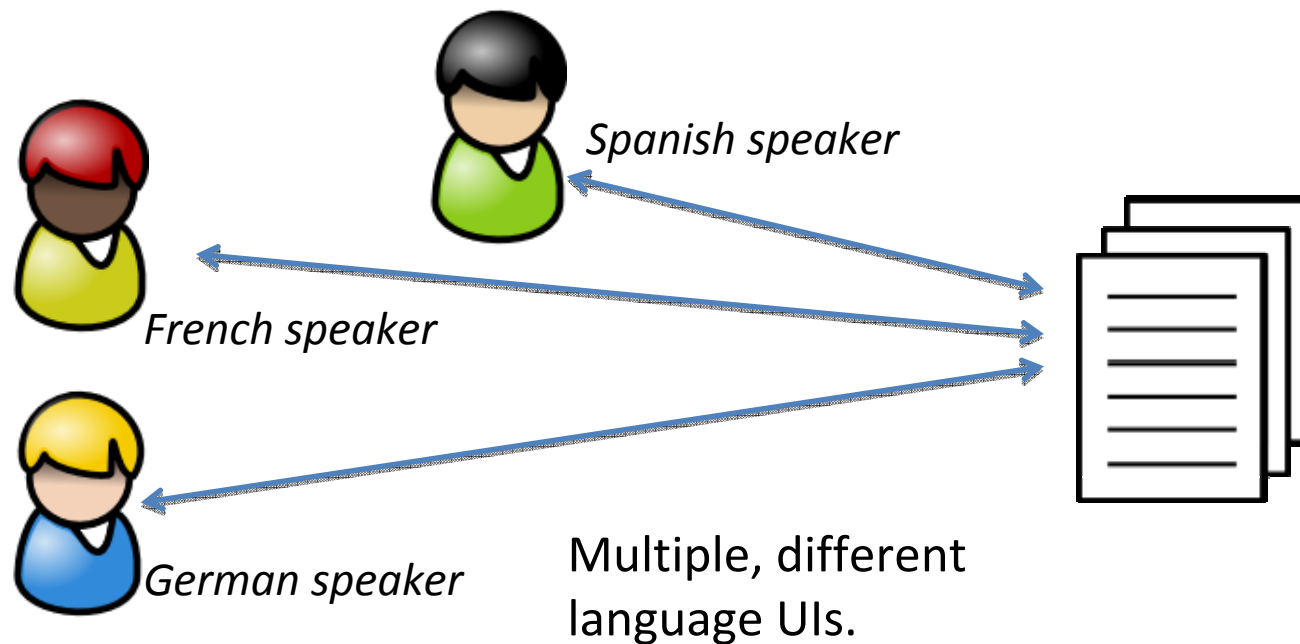
Français

Deutsch

Español

English-speaking user

Taxonomy: Single-language UI. Multiple language translations, not displayed.

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

Spanish speaker

French speaker

German speaker

Multiple, different language UIs.

Bilingual/Multilingual Taxonomies can enable:

3. Different users who speak different languages to search the same body of content that is in multiple languages.
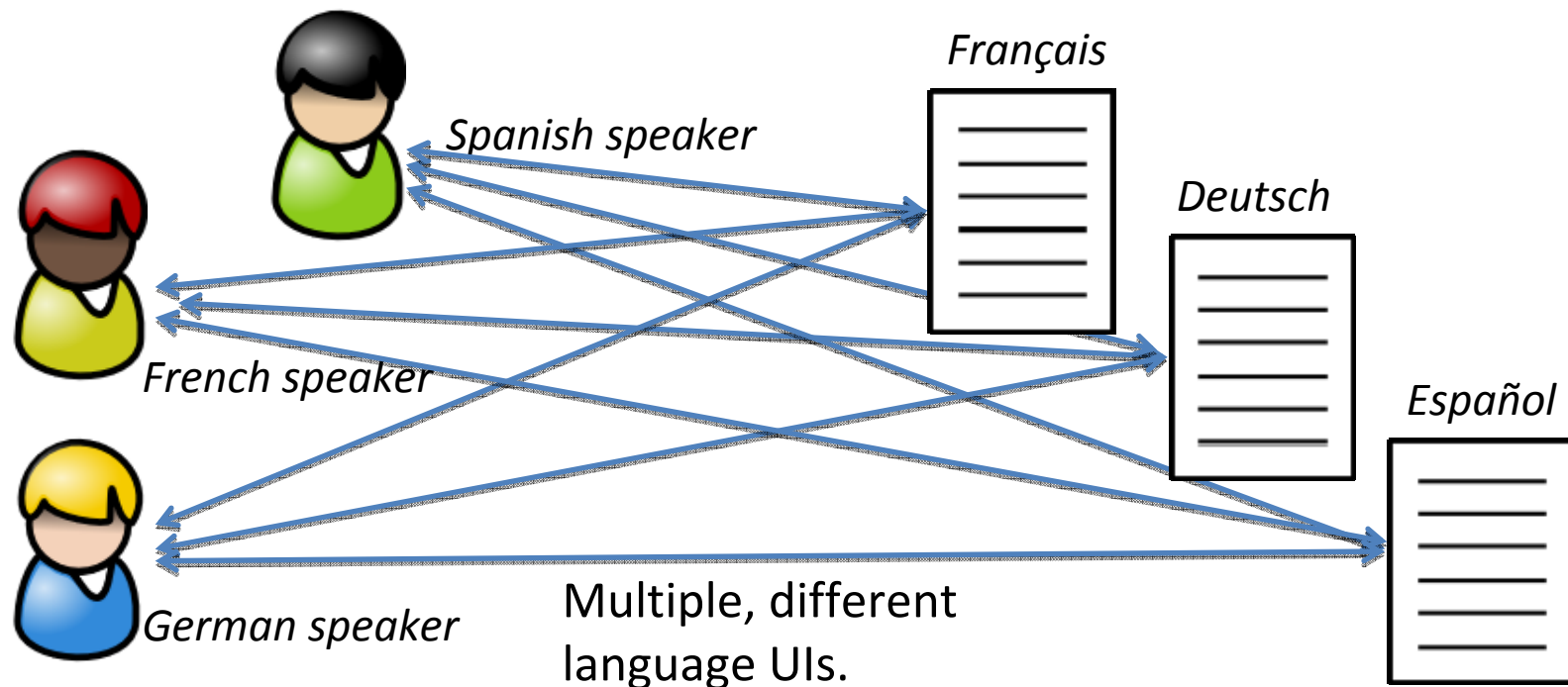
Multilingual Taxonomy Goals

Goals #1 or #2: *Users of one language can access content in a different language.*

- Taxonomy in one language with equivalent translated terms
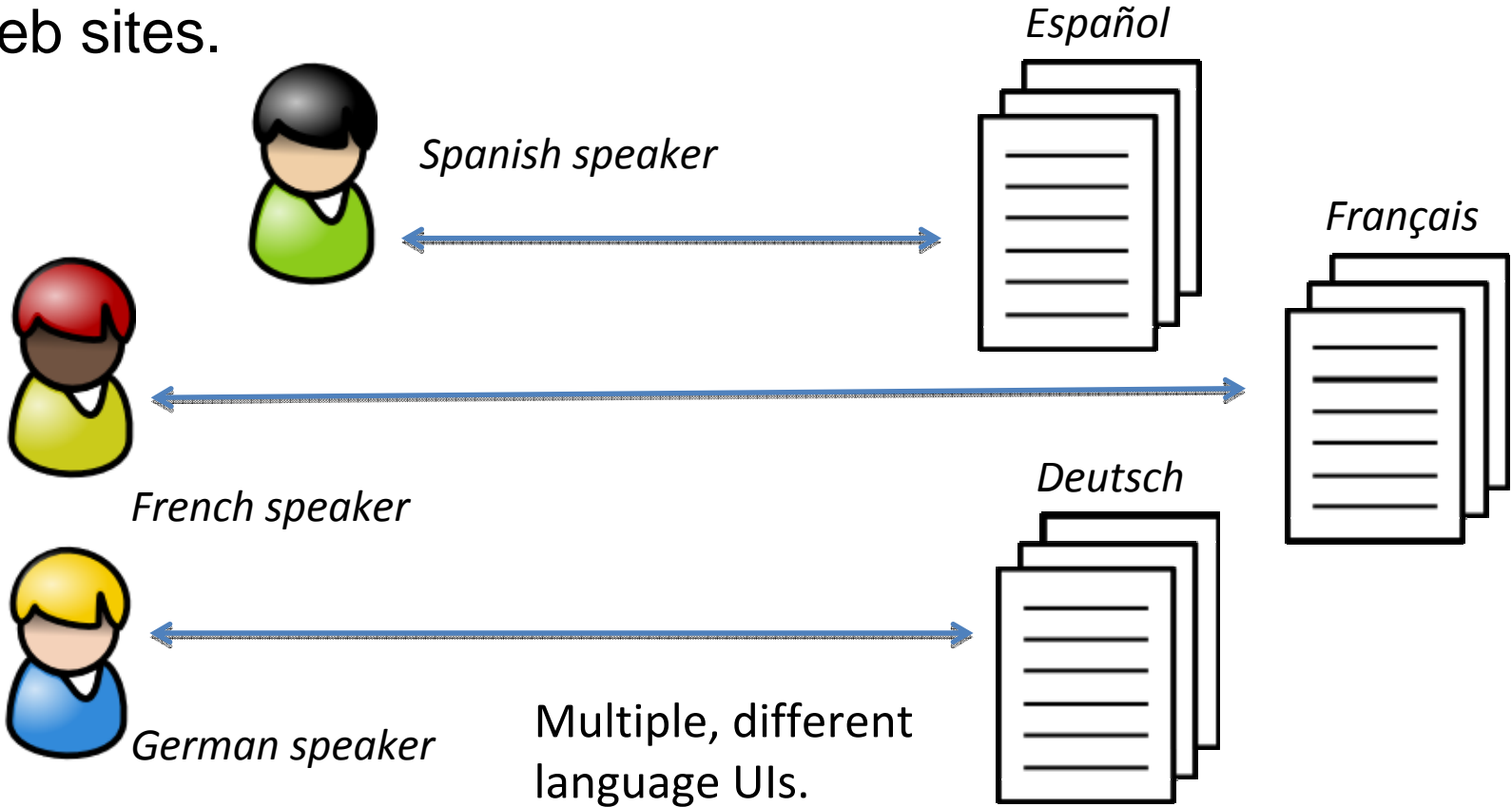- The taxonomy needs to function in only one direction.

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.

Español

Spanish speaker

Français

French speaker

Deutsch

German speaker

Multiple, different language UIs.

Outline

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

Multilingual Taxonomy Design

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)

Multilingual Taxonomy Design

1. Customized field used for term translations
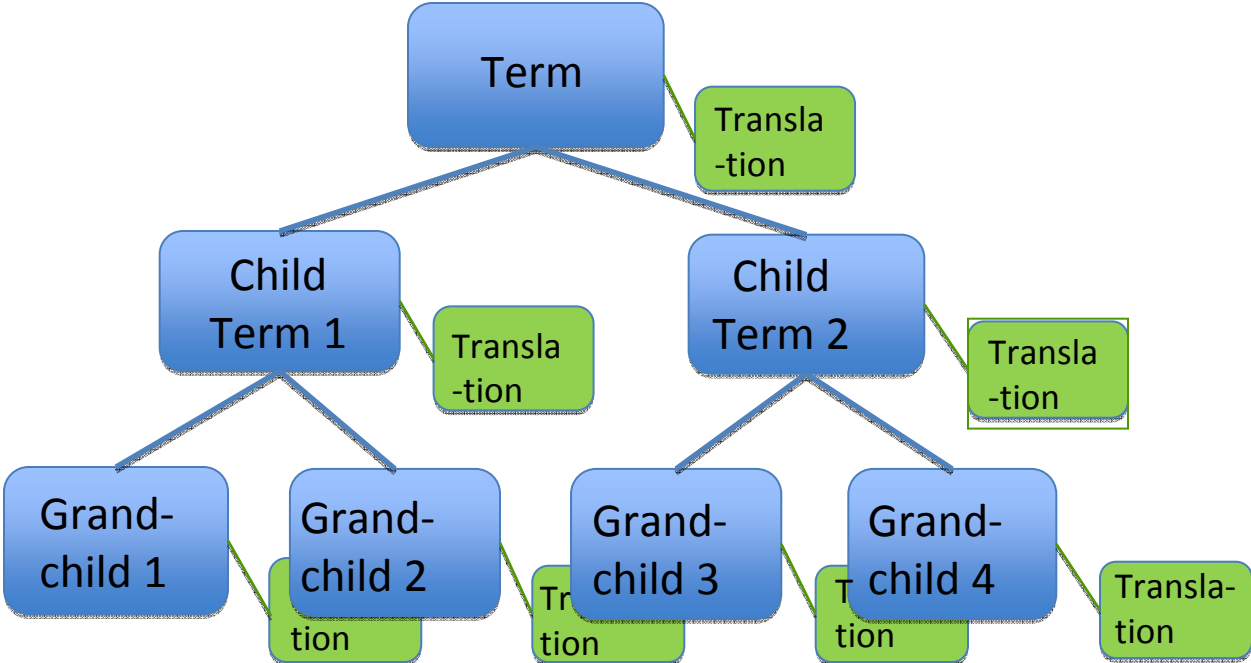
Multilingual Taxonomy Design

2. Translations as a controlled vocabulary flat list, linked at each term

Term

Child Term 1

Child Term 2

Grand-child 1

Grand-child 2

Grand-child 3

Grand-child 4

Term 1 Translation

Term 2 Translation

Term 3 Translation

Term 4 Translation

Term 5 Translation
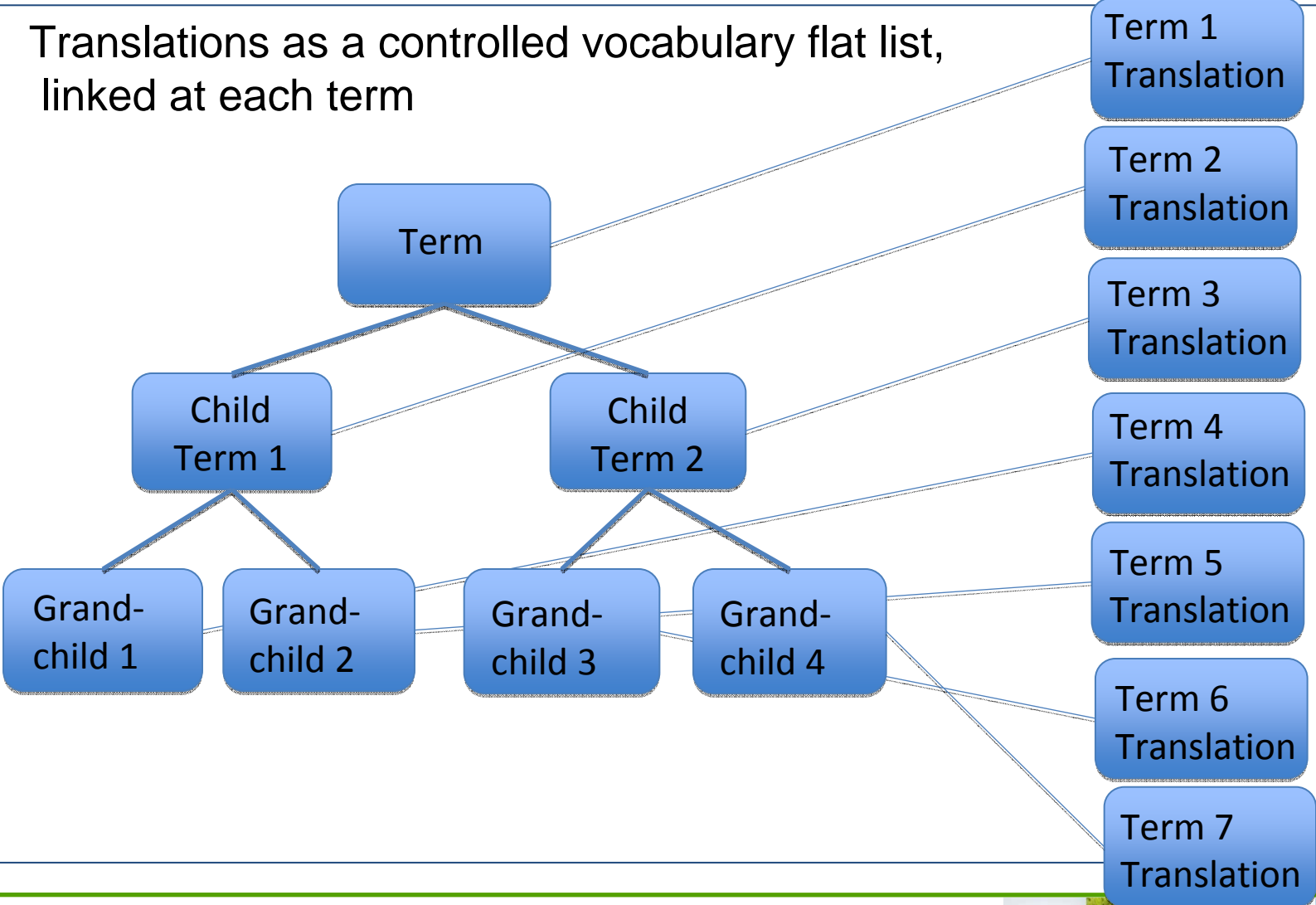
Term 6 Translation

Term 7 Translation

3. Second language taxonomy mirroring first, linked at each translated term. Inter-term relationships replicate.
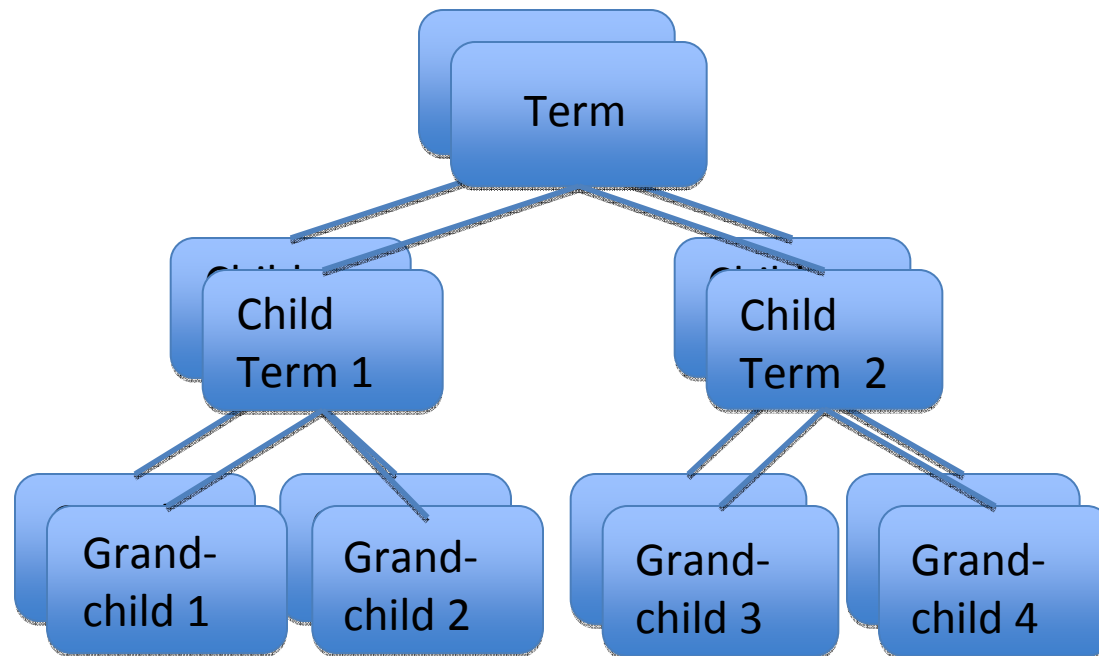
Multilingual Taxonomy Design

4. Multiple taxonomies in different languages linked at equivalent term translations. Inter-term relationships may differ.
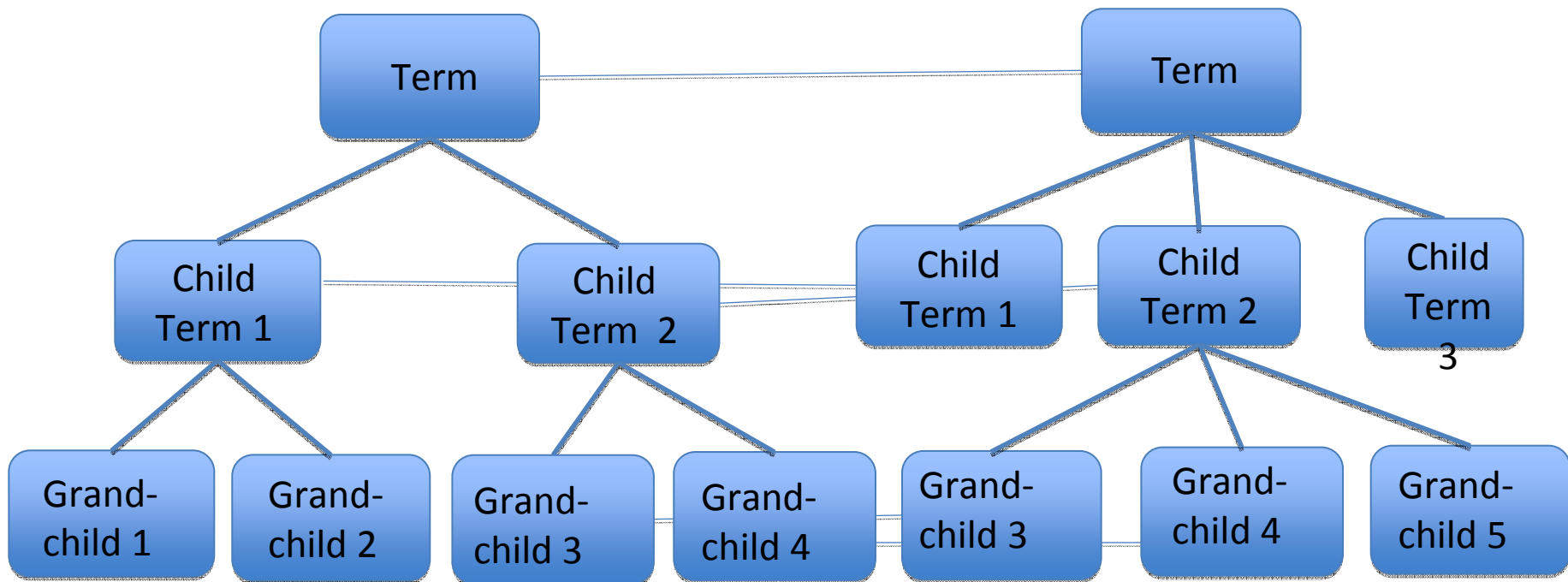
Multilingual Taxonomy Design

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.

Multilingual Taxonomy Design

Method #1:
Create user-defined text field and enter translation

Data Harmony Thesaurus Master

Multilingual Taxonomy Design

Method #1

Synaptica

Multilingual Taxonomy Design

Method #2 or 4: Link equivalent terms in different language by user-defined "alternative language" reciprocal relationship

Data Harmony Thesaurus Master

Multilingual Taxonomy Design

PROJECT performance CORPORATION
Part of the AEA group

Item Summary

Descriptor	Forestry
Object	Test-English
Categories	
Status	Active; Approved; Preferred; Unlocked
UID	1001
Created	dclarke 3/31/2008 12:09:12 PM
Modified	hhedden 11/23/2011 3:45:17 PM

[Save] [Add New] [Subsume]
[Refresh] [Categories] [History]
[Deactivate] [Delete] [Restore]
[Copy]

Sub-Elements

Descriptor:	Forestry
Scope:	

Administrative Attributes:

Approval:	Approved
Workflow:	Normal
Language:	English
Locked:	Unlocked
BLAT:	○ Yes ● No
BTN:	○ Yes ● No

Relationships

[Add/Edit Relationships] [Tree View] [Visualize]

● Limit to Taskview ○ Show All

Forestry

Top Level Parents

TT **Botany** (Test-English)

Parents

BT **Applied botany** (Test-English)

Children

NT **Afforestation** (Test-English)
NT **Reforestation** (Test-English)

Associations

EN>FR **Foresterie** (Test-French)

Variants

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.   *Data Harmony*

Method #3: Create second language taxonomy mirroring first, linked at each translated term. Inter-term relationships replicate.

MultiTes
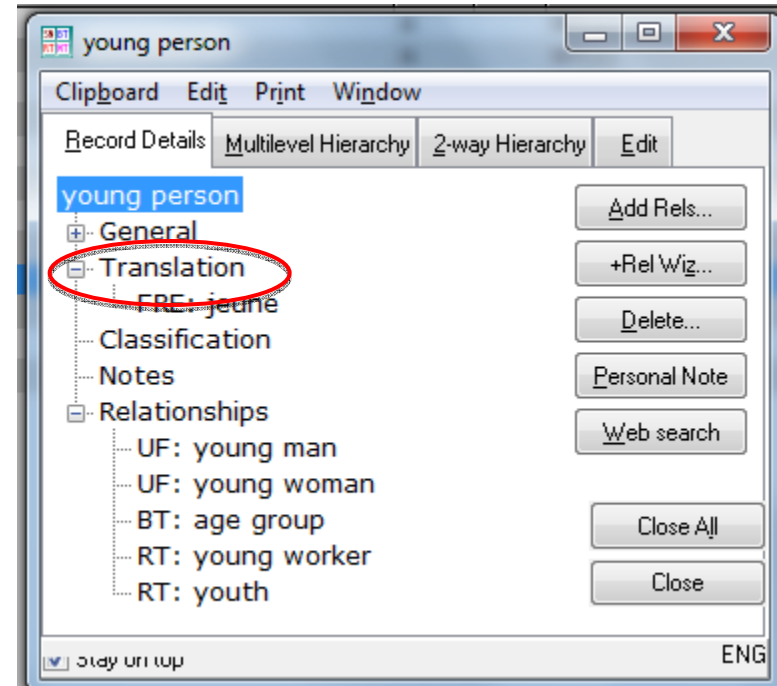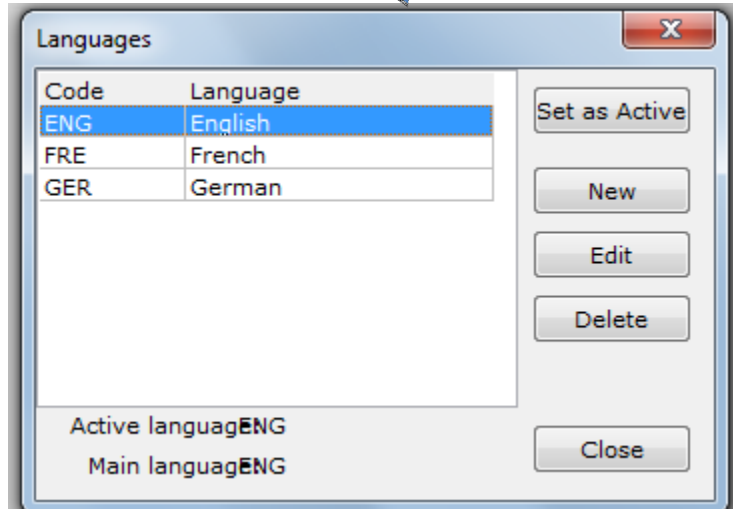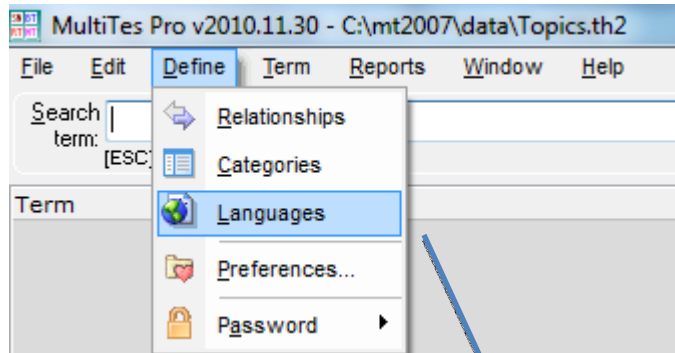
Multilingual Taxonomy Design

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

young person	jeune
FD: 13. Population	**MT:** 13. Population
UF: young man young woman	**EP:** jeune femme jeune homme
BT: age group	**TG:** groupe d'âge
RT: young worker youth	**VA:** jeune travailleur jeunesse
FR: jeune	**EN:** young person

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)

Taxonomy Translating

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

Taxonomy Translating

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.

Contact

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