Sarah Evans

Spring 2006

Thesaurus Evaluation

1) Name and Version/Edition

The European Treasury Browser (ETB) Thesaurus was first created in 2000 and the "final

version" was launched in 2002. Most of the documentation and guidelines refer to the 2002

version though the thesaurus itself has been added to, as evidenced by the addition of five more

languages to the nine languages available previously. I could not find any identification of the

version/edition of the thesaurus I viewed other than "English". I did review the thesaurus in three

different display formats; alphabetical, rotated, and systematic.

2) Features

    a) Source of Terminology

This thesaurus was built on a combination of warrants, though literary warrant was the primary

source. Documents from "European educational repositories, European school curricula,

Socrates project guidelines and education-oriented bibliographic data-bases" were the basis for a

statistical and qualitative analysis to cull terminology. Partly as literary warrant and partly as

cultural warrant, several existing thesauri in the education field were reviewed for input with the

multilingual European Education Thesaurus being the most relevant. In addition, feedback was

solicited from a group of indexers and users that contributed to both terminology and semantic

structure which can be considered user warrant. According to site documents, there are plans for

a feedback form "in order to gather proposals both of new descriptors, non-descriptors, scope

notes, and of deletion or of modification of descriptors or hierarchical relationships" though I

saw no evidence that this had been implemented. There is also discussion of "a more

sophisticated methodology" involving capturing system logs and analyzing search strings when technologically possible. Both of these techniques would increase user warrant in the thesaurus.

b) Specificity of Terminology

My subjective opinion is that this thesaurus is moderately specific in the area of education in general and highly specific in selected core areas. According to site documents, "the scope of the thesaurus is almost boundless" but the thesaurus is "specifically oriented to learning/teaching materials" and "mostly focused on content of formal, informal, non-formal education". In 2002, the thesaurus consisted of 1155 descriptors, 40% of which relate to content. The high number of descriptors in this area infers more specificity.

The level of specificity is also shown within the descriptors themselves. Each descriptor is placed within one of seventeen conceptual areas called microthesauri which have additional levels of specificity notated by numbers. Narrower terms (NT) and broader terms (BT) are each given a number to show their level in the hierarchy. Most NT are on the first level but a few, particularly in the area of content, are as specific as level 4. The following is an example from the conceptual area CONTENT (70) under the heading of HUMANITIES (70.20).

literature (descriptor)

. . . . NT1 non-fiction
. . . . . . NT2 biography
. . . . NT3 autobiography
. . . . NT4 <u>life story</u>

c) Semantic Relationships Shown

The structure of the thesaurus is described as the "classical semantic relationships in a thesaurus." As described in section b, descriptors are classified into 17 conceptual levels called

microthesauri and further divided into sub-groups. Within the thesaurus are examples of the three major types of subject relationships, equivalence, hierarchical, and associative.

Equivalence Relationships

The ETB thesaurus has two specific kinds of equivalence relationships. Inter-language equivalence is a priority because of the purposes of the thesaurus. Each term in one language version has an equivalent version in each of the other languages. It is designated by a two letter code representing the language. For example:

mammal
Da: pattedyr
De: Säugetier
El: θηλαστικό
Es: mamífero
Fi: nisäkäs
Fr: mammifère
He: יונק
Hu: emlõsállat
It: mammifero
Nl: zoogdier
Sv: däggdjur

Intra-language equivalence is demonstrated with the standard USE and UF (use for) designations. Non-preferred terms are in italics. Two examples are below.

manual communication
USE nonverbal communication

maladjusted pupil
UF *special needs pupil*

It is notable that a significant number of non-preferred descriptors have been encouraged in each language "in order to increase the access key to semantic information" since end-users language choices do not always match the indexers.

Hierarchical Relationships

Each descriptor is linked to other descriptors in a "mutual and asymmetric relationship." A

descriptor can have more than one broad term (BT) and, as described in section b, narrow terms

(NT) have an exponent that indicates its level in the hierarchical chain. Both genus-species

relationships, here called "instance relationships" and whole-part relationships are considered

hierarchical in this thesaurus. This example of a genus-species relationship comes from the

MATHEMATICS AND NATURAL SCIENCES (70.40) sub-group of the CONTENT OF

EDUCATION (70) microthesaurus:

> sciences
> . . NT1 mathematics
> . . . . NT2 algebra
> . . . . NT2 applied mathematics
> . . . . NT2 arithmetic
> . . . . NT2 calculation
> . . . . NT2 geometry
> . . . . . . NT3 analytical geometry
> . . . . NT2 mathematical analysis
> . . . . NT2 statistics
> . . . . NT2 trigonometry

This example of a whole-part relationship comes from the EQUIPMENT (80.20) sub-group of

the FACILITIES/EQUIPMENT/MATERIALS (80) microthesaurus.

> equipment
> . . NT1 computer
> . . . . NT2 hardware
> . . . . NT2 software

> Associative Relationships

The guidelines for associative relationships in the ETB thesaurus are described as "a relationship

between terms which are not members of an equivalence set, neither are subordinated or

superordinated to another, but are mentally associated to such an extent that it's useful to make

the link between them explicit." In addition to the designation RT, a number is given for

reference if the descriptor is classified under another microthesaurus and subgroup. Also, related

terms may have related terms listed as well. This kind of organization gives contextual richness

to the relationships. The following example comes from the microthesaurus INDIVIDUAL DEVELOPMENT (10).

> life cycle
> . . NT1 adolescence
> . . . . RT adolescent
> . . . . RT puberty
> . . . . . . RT sex education (70.70)

In this case the related term "sex education" can be found in the microthesaurus CONTENT OF EDUCATION (70) in the subgroup CROSS-CURRICULAR/EXTRA-CURRICULAR ACTIVITIES (70.70).

d) Term Syntax and Semantics

I could not locate explicit guidelines in the web documents for most syntax or semantic rules other than the statement that this thesaurus conforms to ISO standards. For example I could not find guidelines for singular/plural forms or homophones. Based on my review of the thesaurus, terms are not inverted and are entered directly. The examples given I've given in the previous sections support this. There are scope notes scattered throughout the thesaurus for a few select terms.

e) Pre-coordination vs. Post-coordination

The ETB thesaurus is described as a "flexible thesaurus for a modular post-coordinate retrieval." Indeed, the majority of the terms reflect a single concept such as "athletics" or "orienteering." Where a term includes a preposition or conjunction it is typically still a description of a single concept such as "education for peace."

f) Creation of Terms

This thesaurus is decidedly an enumerated vocabulary that is technically closed. It could be considered partially open in the same way the ERIC Thesaurus is open. As described in section

a, the creators of this resource intend to accept input on term suggestions from the greater community and, eventually, monitor usage logs for search strings that could become new descriptors.

g) Other Noteworthy Features

The thesaurus vocabulary is available to be displayed in three ways, alphabetical, rotated, and systematic. In the alphabetical display, "each descriptors can be displayed accompanied by the net of its semantic relationships, so that alternative terms for indexing by or searching for are suggested." This is also allows you to view the inter-language equivalencies. For the rotated display "terms are displayed in alphabetical order, in such a way that compound terms comes more than once in the list, having each significant word as a word of order." For example:

creative **activities**
extracurricular **activities**
hands-on **activities**
leisure **activities**
leisure time **activities** USE leisure activities
outdoor **activities**
school **activities**

I personally did not find the rotated display as useful, but I imagine it could be if involved in an indexing project. The systematic display allows you to view the terms in their microthesauri and subgroups with the hierarchical relationships.