

# thesauri in libraries

LIBR 506  
Week 7 Feb 28 2022

## 1 Lecture

What principles do we use to determine what terms are appropriate and need to be included?

How does this play out in the Library of Congress Subject Headings?

### 1.1 What are Thesauri?

**Thesauri:** we encounter these often in print materials and e-books, in articles

Note: an author-supplied list of keywords isn't considered a thesaurus (not a controlled vocabulary, not selecting from a limited or controlled list of options)

Whereas a list of subject keywords applied by a professional indexer is an example of a controlled vocabulary

### 1.2 "Aboutness"

**Aboutness:** "the property of a word", or "how well a word indicates its subject matter"

- How do we make a word stand in for a concept in our minds?
- Some terms have not much aboutness (e.g., "the"), and some have a lot (very specific to what we expect to find in a resource—e.g., medical terms)
- How can we systematically pick words with the right aboutness to represent the concept in our minds, the resource? 3 big phenomena:
  1. **Evidence** (what we will focus on in this lecture)
  2. Epistemology (how can I be sure that the way I am thinking of a term and an associated concept in the same way that you will?—How do I know 2 people have the same word for a concept?)
  3. Obsolescence: will the concept be a good representative over time? (e.g., trends and needs will change over time, terms themselves will fall out of favour, etc.)

### 1.3 Evidence (Warrant)

**Evidence (Warrant—**used in information organizations):

- Warrant can help us determine the **right** (preferred, stable) word for a concept
- Help us determine which words are **equivalent** in meaning (e.g., lead-in terms)

- Can help us determine how words are **related**

Some common types of warrant:

### 1.3.1 Literary Warrant

Both the oldest and most established type of warrant for knowledge organization systems

Uses **“the literature”** as a source of valid terminology

- Could be the collection itself, or the full domain of knowledge that your collection represents

Determines what terms are used (preferred terms) and the level of specificity for different subjects

Philosophy behind literary warrant: you are presenting the collection in the collection’s own voice; extracting terms from the resource (or the discourse, or the scholarly community it is taking part in)

Leads to decisions like:

- “Gold” is not a valid subject heading (because there are no books on it), even if other elements have subject headings
- You have to set thresholds of how many times a concept must occur in new items before a subject heading is added
- Have to decide whether to re-index existing items to match new items, or split the records (e.g., because of language shifts, representation issues—do you go back and fix that, or say “these works really are about this older term”?)

### 1.3.2 Scientific (consensus) Warrant

Appealing to the people and the experts to determine the valid terminology

“Trying to create a system that reflects reality itself” (very Enlightenment idea)

Usually makes the collection conform to the agreed-upon terms within a community of expert; ideally creates a system that reflects fields of knowledge accurately

As a user learns to use a system, they will also learn the logic of that scientific community

Leads to decisions like:

- “Gold” will be a valid subject heading because it is on the periodic table
- Need to make decisions sometimes about which “camp” in a scholarly community has the authoritative view of the discipline (e.g., “library and information studies” or “information science”)
- Need to decide how to update terminology and term relationships when there are major shifts in a scientific paradigm

### 1.3.3 User warrant

**“Users”** are a valid source of valid terminology: you are trying to get to the minds of your users and the terminology they will expect as the valid terminology in that system

For example:

- Look at search logs: did they find the resource they were looking for?
- Or developed directly by surveys, interviews

- Most powerful in creating exhaustive lead-in terms (e.g., you could augment literary or scientific warrant with user warrant to make lead-in terms to help guide users)
- Ideally a user warrant system creates an intuitive system of users' natural language, have fewer mediated steps between the term in the user's head and the source they are looking for

Leads to situations like:

- Should the system represent an "average" user or a multiplicity of user types? Need to decide who you are trying to reflect in the system
- You will end up using novices' terms for items and concepts, rather than terms in the items themselves (could be inaccurate, flawed)
- You will have to set thresholds for how many times a new term appears in search logs before you create a subject headings

### 1.3.4 Ethical warrant

Usually not the premise of a system as a whole, but a corrective step to make sure the system isn't causing harm

For example: making corrections or adjustments to consider religious, racial, cultural, gender, language representation

E.g., correcting for bias, removing slurs

Leads to decisions like:

- Revising terms from slurs used by the majority to respectful terms used by the community itself
- Might create symmetry in subject headings to overcome bias (e.g., correct an unmarked category "contraception" by adding "female contraception" in addition to "male contraception")
- May want to create the same level of detail regardless of the proportionality of holdings (e.g., religions)

## 1.4 Library of Congress Subject Headings

Key features of the Library of Congress Subject Headings: (LCSH)

- In nearly every English book you'll come across
- Not quite a thesaurus
- There is a set process for how to propose revisions to terminology (collaborative process by experts in DC)
- It has subdivisions:
  - Principle of pre-coordination: putting together many representative terms of a work in a string. (e.g., by geography, chronology, form, subdivision by other types of topics, multiple subdivisions)

LCSH is not quite a thesaurus because of the "hierarchies" it contains:

V History

V Historic sites

V Memorials

V War memorials

V Military trophies

(are "historic sites" really a "part / instance / type" of history?)

Do "historic sites" inherit all the properties of "history"?

### 1.4.1 Growth & Change in the LCSH

The LCSH headings are applied to enormous libraries of works; meant to be a universal system that works for all human knowledge, therefore constantly needs to be updated

If you want to propose a new term in LCSH, typically happens through **cataloguing-in-publication**: librarians get access to the text of a book before it gets published, will interpret the work according to existing terms, and if there is no subject term that adequately covers, they will propose a new term

Website about the process: <https://www.loc.gov/aba/cataloging/subject/lcsh-process.html>

And monthly reports: <https://www.loc.gov/aba/pcc/saco/cpsaed/cpsaeditorial.html>

Cataloguing lab (place for library professionals to discuss subject headings to put the best one forward): <https://cataloginglab.org/kbtopic/subjects/>

“Illegal Aliens” in the LCSH: how the process was supposed to work and didn’t

- Proposal to cancel **illegal aliens** and instate **noncitizens** and **unauthorized immigration** instead: <https://www.loc.gov/aba/pcc/saco/cpsaed/psd-160321.html>
- Report of the Subject Analysis Committee Working Group on Alternatives to LCSH “Illegal aliens”: [https://alair.ala.org/bitstream/handle/11213/14582/SAC20-AC\\_report\\_SAC-Working-Group-on-Alternatives-to-LCSH-Illegal-aliens.pdf](https://alair.ala.org/bitstream/handle/11213/14582/SAC20-AC_report_SAC-Working-Group-on-Alternatives-to-LCSH-Illegal-aliens.pdf)
- Term has gone from being in limbo to one that no one is happy with

What competing principles or warrants are at play here? What is the relevance of the political and institutional context to the design, maintenance, and revision of LCSH?

Looking at controlled vocabularies, you can consider starting from the perspective of warrant what the system is supposed to do:

- What is the system’s relationship to reality? To the state of knowledge? To how we wish the world to be?
- How is the system meant to serve the users? Authors of the collection? Society?

## 2 Readings

**2.1 Library of Congress Cataloguing & Acquisitions. Process for Adding and Revising Library of Congress Subject Headings**  
<https://www.loc.gov/aba/cataloging/subject/lcsh-process.html>

**2.2 Bone, C., & Lougheed, B. (2018). Library of Congress Subject Headings Related to Indigenous Peoples: Changing LCSH for Use in a Canadian Archival Context. Cataloging & Classification Quarterly, 56(1), 83-95.**