Library of Congress Subject Headings

Module 1.4:
How Do We Determine Aboutness?

Policy and Standards Division
Library of Congress
June 2016

Subject Analysis Process

- Conceptual analysis of item

- Translation of the aboutness into specific notation or terms

In the previous module, we discussed *why* it is important to analyze a resource to determine what it is about, but we didn't explain how it is done. So, how do we perform subject analysis?

Well, the problem is, I can't tell you exactly how it is done, because no one agrees completely on the number of steps involved or the nature of those steps.

But, generally, there are two activities that are considered to be essential:

First, we have to examine the resource and figure out what it is about. In other words, there is a conceptual analysis.

Then we have to translate that aboutness into one or more subject languages, such as *Library of Congress Subject Headings* (LCSH*), Art and Architecture Thesaurus* (AAT), Library of Congress Classification, Dewey Decimal Classification (DDC), and so forth.

Others have included additional steps in their analysis, such as:
- prioritizing the concepts and
- familiarizing yourself with documentary languages,
- sketching out a classification hierarchy,
- picking words to search in the vocabulary,
- and the like.

Of these two essential steps of determining aboutness and translating that aboutness, we have a basic understanding of how that translation is done. Most of our subject analysis literature has focused on that process.

Principles and Concerns

- Specificity
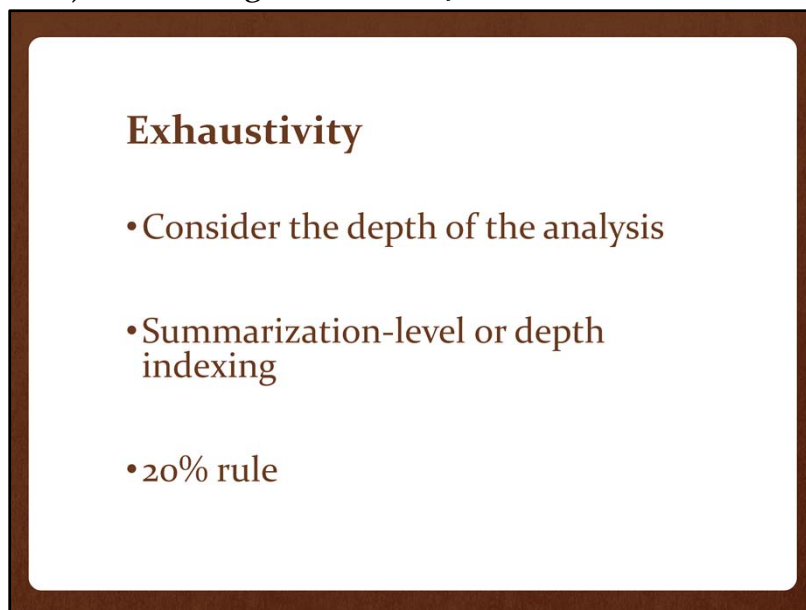
- Objectivity

- Exhaustivity
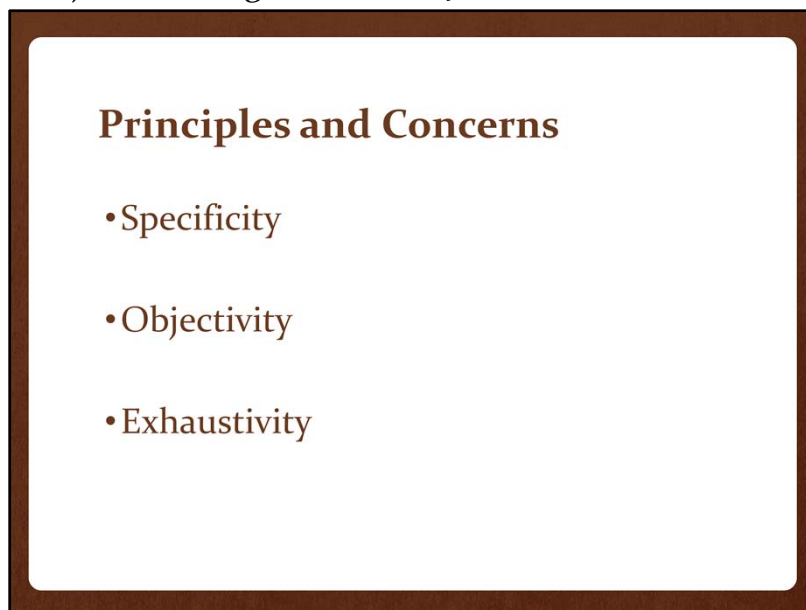
When thinking about subject analysis, we take into consideration some underlying principles and concerns. One of the most important principles we deal with is the idea of *specificity*. We are specific in subject analysis in how we describe the item's aboutness. We're specific in the terminology that we choose to represent concepts. We want to get at the essence of the topic.

Despite the fact that subject analysis can be a highly subjective process, it's a very interpretive kind of activity, and we recognize that. We want to make sure that in this process we try to remain as objective as possible despite the fact that we know that the process can be culturally biased – that we are influenced by our own culture, by our own point of view, from our own experiences. This is how we see the world.

We also recognize that our controlled vocabulary systems and our classifications are culturally fixed. They are from a place; they are from a time; they do reflect a point of view. By realizing that and accepting that, we try to remain neutral, suspend disbelief, accept the idea that all knowledge is equal, and attempt to be as objective as possible.

In this process, self-awareness is key. We need to know what kinds of issues push our buttons. We need to know when to let someone else analyze the material instead of us. And we must remember that our ethical behavior is relied upon. Catalogers hold tremendous power. A cataloger can make any resource in the collection "disappear" with simply a little typo here, the wrong subject heading there, maybe transposing digits in a classification number… But we know that our users rely on us to make sure that all materials, no matter how we personally feel about them, are accessible to all.

Exhaustivity

- Consider the depth of the analysis

- Summarization-level or depth indexing

- 20% rule

While performing subject analysis, one of the things we must consider is the level of *exhaustivity*.

Exhaustivity is the level at which the analysis is going to take place; it addresses the number of concepts that one is going to include in the aboutness determination and the translation stages of subject analysis. When we think of exhaustivity we need to think of it as a continuum, from summarization to depth indexing.

Summarization is where the cataloger identifies only the dominant, overall subjects of a resource. We recognize only the concepts or the main theme of the material. This is the traditional approach that we tend to use with bibliographic records in library catalogs. We look for the overall subject that is the topic of the entire information resource.

Depth indexing, on the other hand, looks at extracting all of the main concepts of a resource, but also many subtopics and sub-themes. If one is doing depth indexing, then many topics covered only in small parts of the resource will be recognized. When we think about depth indexing, one of the things that might help clarify what we mean by this, is to think of a search engine as being the ultimate in-depth indexing, because all of the words in a document are included as index entries.

It is important to remember that the level of exhaustivity may be determined on the local level. An institution may have a policy as to the number of subject headings, or the number of concepts that may be included in the bibliographic record.

Conceptual Analysis

- What is it?

- How is it done?

Many people want to know, what *is* the activity known as conceptual analysis?

I was disappointed when I first became involved in the library and information science professions, to realize that catalogers did not sit in a back room somewhere reading every resource that crossed their desk before they cataloged them. Well, I quickly learned that that is not realistic. That is not a possibility. And I also learned that there are a lot of resources out there that I don't want to read!

Catalogers don't have the time to read every resource, word-for-word. There are too many items to deal with. There aren't nearly that many catalogers anyway. It's not a practical, economic way to approach technical services work.
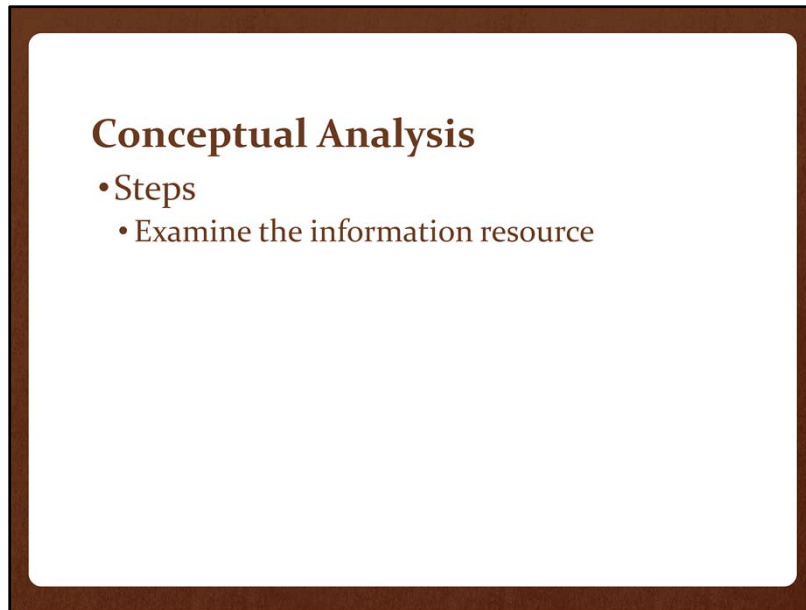
So, some sort of short-cut has to have been developed. So what is it?

Well, there really isn't one way of doing this. The best approach, that many textbooks have come up with, is just to list some bibliographic features for the cataloger to look at. But they never really tell you exactly how to make the determination of what something is about.

We know what the process is supposed to do: Conceptual analysis is an attempt by an indexer or cataloger to determine the aboutness of the document. And we can re-phrase that over and over again in any number of ways.

- We can talk about discovering the nature of the document.
- We can look at what the document covers.
- We discern the topics discussed in the resource.
- We find the indexable matter in the document.

We know what we want from Conceptual Analysis – an understanding of what something is about – but little has actually been said about how it is done.

Conceptual Analysis
- Steps
  - Examine the information resource

In this next section, we will try to outline some activities for you to perform as you do conceptual analysis.

And we will look at some models of analysis that have been discussed.

The first step is to look at the resource.

Examining the Resource

- Examine the parts of the resource that stand out.
  - titles
  - table of contents, chapter headings, sections
  - preface, introduction or its equivalent
  - illustrations, diagrams, tables, captions
  - conclusions
  - others

We'll start by examining the parts of the resource. And at this point it's a really good idea to start writing things down. Writing things down as you see them is essential in order to produce a clear picture and not some hazy outline.

So what are we looking for? We're looking at the titles, the subtitles, series titles, and other sorts of titles. Some titles are good, and, as we know, some are bad for determining aboutness. Some can be ironic, absurd, comic, and so forth.
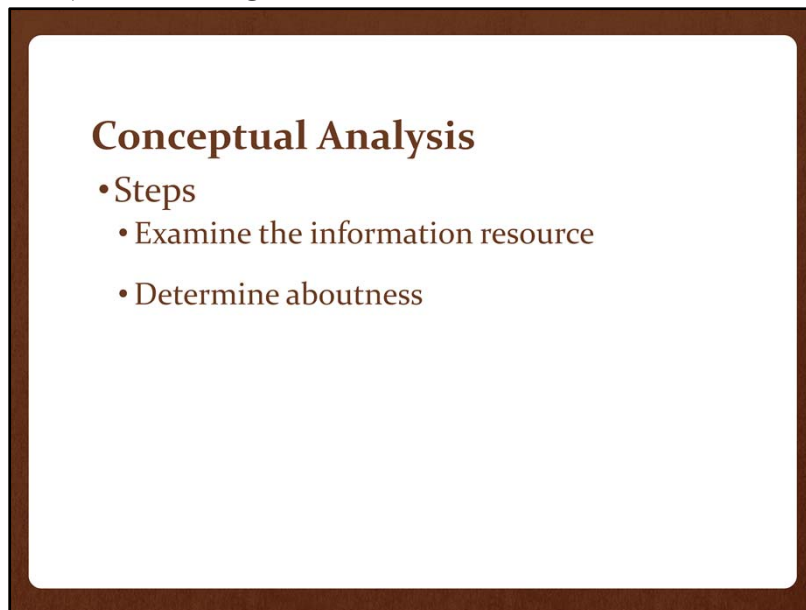
Knowing who the author is might help, too! Perhaps the authors provide their field of study in the resources themselves. Or perhaps we go look them up and determine what else they've written. That might help, too. We also can examine the table of contents, the chapter headings, section headings, appendices, and their equivalents.

Now we want to examine the preface or the introduction, and it's a very good idea to read this. This is an excellent source for learning what the author's intent is. We're looking for the thesis statement, the plan of the resource, the premise, and the author's point of view.

If the work is textual, we can also look for headings that are in bold, or text that is in bold, which helps us know what the author and the publisher believe to be important. Are there illustrations? Are there captions? Containers, dust jackets, site maps, credits, and the information in them can also be very important.

If, after doing all of that, you're still not sure what that resource is about, read a portion of the text, watch a part of the video, whatever it might be. Perhaps in some cases the resource is short enough that you *can* read the whole thing.

If all else fails, check out book reviews, because they are going to help you know what other people thought that that resource was about.

**Conceptual Analysis**
- Steps
    - Examine the information resource
    - Determine aboutness

As we've said, there is no single way to determine what something is about.

Each resource that you examine might require different methods.

Some materials may be complex; and they may require multiple methods to determine what they are about.

Catalogers have to determine what methods work best for them. In the next few slides, we'll review some of the methods that have been discussed in the library and information sciences literature.

Library of Congress Subject Headings: Module 1.4
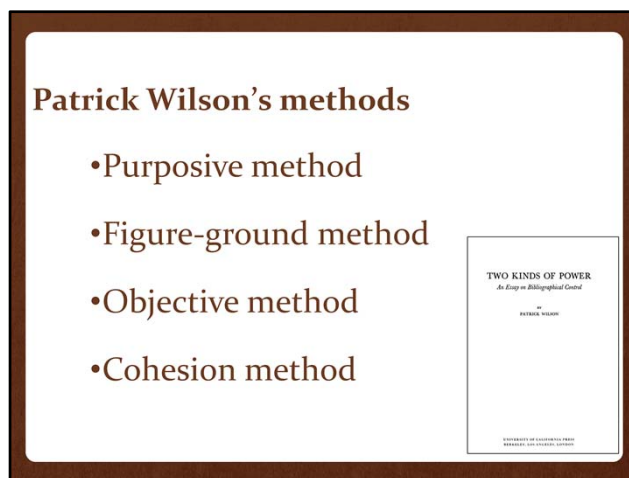
Patrick Wilson, a philosopher who taught Library and Information Science at UC Berkeley, wrote about some ways of carrying out subject analysis. He wrote about this in his seminal work, *Two Kinds of Power: An Essay on Bibliographic Control*. In this work, one of the points that Wilson brings out is the idea that there is not really one obvious subject for any given resource. Something that can be labeled "the subject" might be an expectation on the part of catalogers and readers, but the actual subject matter of resources is more complex than that. It goes beyond just "the facts", or the statements found in the resource itself. There are also relationships between ideas. There are influences. There are biases. There are things that are left unsaid. There are *implied* ideas. All of this can create a level of complexity that one might not expect when one talks about finding "the subject".

Wilson offered four approaches for this subject analysis process. He was not convinced that any one of them was better than the other. He did not see that there was a right or wrong way to determine aboutness. It may be one of the four approaches he came up with – it may be none of them. It could be more than one – you could use all of them at different times. Or it could be a few. It really depends on the resource that you are working with and the person who is doing the analysis to determine what processes work best for them.

Now the methods that Wilson offered include what we refer to as the **purposive method**. This is the idea that you will look at what the creator's aim or purpose is, to help determine what the resource is about. The problem with the purposive method is that it's not always clear. An author may say once what their aim or purpose was. A creator may say it multiple times. Sometimes the multiple times is less helpful because the aim or purpose might contradict, or conflict with other statements. It might be said not at all. So in some ways we might be guessing as to the author's aim or purpose. With non-fiction materials, it can be fairly straightforward at times to use the purposive method, but when one starts analyzing literature, the arts, the humanities, it becomes a little more complex. It becomes a little less clear as to what someone's purpose might have been in creating that work.

Another method that he offers is the **figure-ground method**. This is the idea that when looking at a resource, *something* is going to stand out. It might be an idea; it might be a historical figure or a character; it might be a particular concept or philosophy. *Something* will stand out. But the problem with the figure-ground method is the idea that what stands out to one person, may not stand out to another. In other words, the reader, the analyst, also has an influence in interpreting what something is about. So, what stands out to me one day, may not stand out to me three years later, if I go back and revisit that resource.

These are the first two of Wilson's methods.

Patrick Wilson's methods

- Purposive method
- Figure-ground method
- Objective method
- Cohesion method

Wilson's third method was the **objective method**. And this was the idea that we could count the number of references to ideas or subjects that appear in a resource. If we look for topics or subjects that are mentioned most throughout, perhaps that is what the work is about. This may be helpful in the determination of aboutness, but there are also some issues with this approach as well.

For example, what may be mentioned most may be a background idea, or a setting for the main topic. Or, in another case, perhaps the major topic is never explicitly stated. Maybe it is implied in the work and it might be overlooked. Also, something may be named the first time it is mentioned in the resource, but there may be synonyms used to describe it later. If we are going to use something like the objective method, we need to be able to cluster ideas and concepts. It is one of the challenges of this to make sure that we see the connections and the relationships. A simple, straightforward, objective counting of references might actually lose the relationships between various concepts if they are described using various words or names.

The fourth method that Wilson offered we refer to as the **cohesion method**. And this is the idea that maybe we can determine the aboutness by looking at what holds the work together. What is the central idea? What is the central theme? What is the cohesiveness there? What has been said; what has not been said?

The problem with this approach of looking for what holds the work together is that it requires a certain amount of subject expertise to know what's there and what's been left out. Determining what something *is not* is not an easy way of determining what something *is*.

Langridge's Conceptual Analysis

- What is it about?

- What is it?
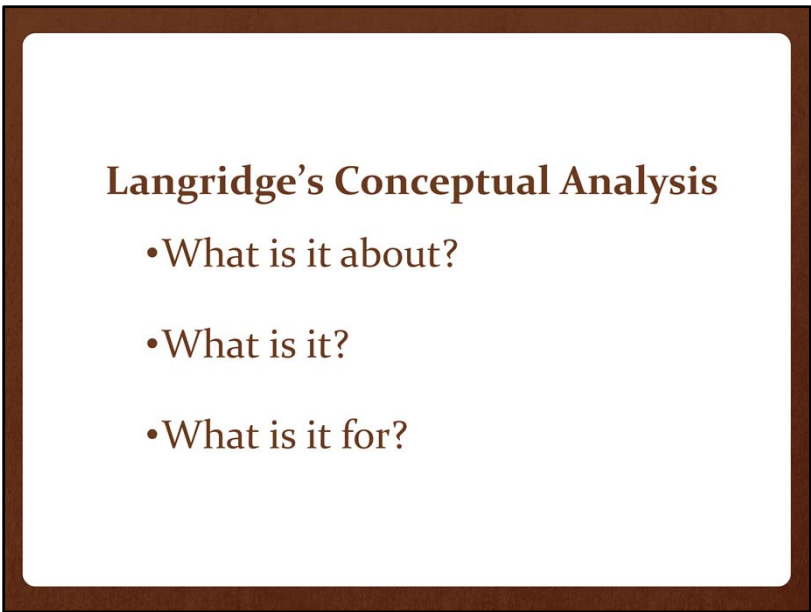
- What is it for?

Derek Langridge, another LIS scholar, takes a different view of the conceptual analysis process. He thinks you need to ask the right questions and those questions are what you see on the slide.

What is it about? That addresses the topical aspects of the work, but also, he wants you to look at:
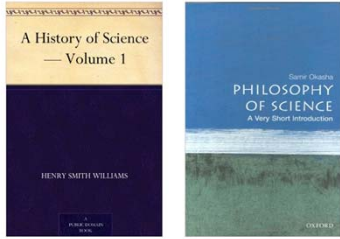
What is it? He's looking at the *form* of knowledge. And he also says you should be asking:

What is it for? This is a question that addresses the discipline of a work, the uses of a document, and maybe to whom it is addressed. Its answer, what is it for, may particularly affect the choice of classification. A book about anatomy for nurses is not the same thing as a book about anatomy for veterinarians.

Langridge thinks aboutness only gets confusing when we mix up these three questions. This is what leads to variations in our impressions of aboutness.

To understand Langridge's point of view, let's look at these two titles.
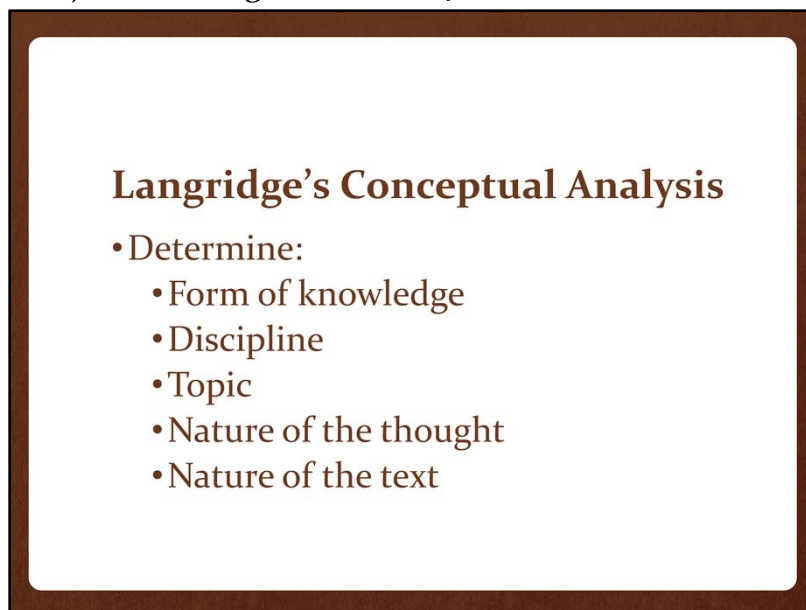
Here we have two works. One is a *History of Science*. The other is *Philosophy of Science*.

If I ask, **"What are they about?"** you'd probably say, "Well, they're both about science." Yet, looking at the contents of each of these resources, we'll find that the material is vastly different. The subject can have more than one meaning – can have more than one perspective.

You could get more specific, though. When we look at "What is it?" and "What is it for?" we bring in these clarifying points. The *History of Science* is looking at the past activities of people calling themselves scientists, who pursue the knowledge of nature. The *Philosophy of Science* addresses a system of beliefs accepted as authoritative in science. These are vastly different kinds of works because they are in two different forms.

"What is it?" – that question – it clarifies it. "What is it?" *History of Science* is a work of history. *Philosophy of Science* – obviously, a work of philosophy.

"What is it for?" As I mentioned earlier, this can have a number of answers and it can relate to the audience for the work, or the discipline in which it belongs, or the various uses for that document.

In addition to the questions that Langridge wants a cataloger to ask, he also believes that the process of determining aboutness involves an examination of these five components that are on the screen.

The first is the **form of knowledge**. Langridge wants to see what the form of the resource takes. In his own research he identified twelve fundamental forms of knowledge, including things like philosophy, history, art, religion, and so on.

He also believes that identifying a specific **discipline** is important. Disciplines fall under those twelve forms of knowledge. Disciplines are also the basis for most of our classification systems. So, choosing the discipline in which a resource falls, can be very helpful in the classification process.

Topics. Of course we need to determine what the topics of a resource are. We want to look at what the subject matter is, and there are billions of possibilities there.

According to Langridge, we also look at the **nature of the thought** and the **nature of the text**. Many LIS scholars might not mention these as components of the aboutness determination process, but Langridge felt that we might get a better sense of the resource itself, if we understand the creator's point of view, the culture in which the work is situated, the audience, and the intellectual level of the resource. In terms of the nature of the text, he felt that perhaps the type of writing, or the medium of communication, or the structure of the work may also be important in determining what the work is about and how we ultimately describe them in our bibliographic descriptions.
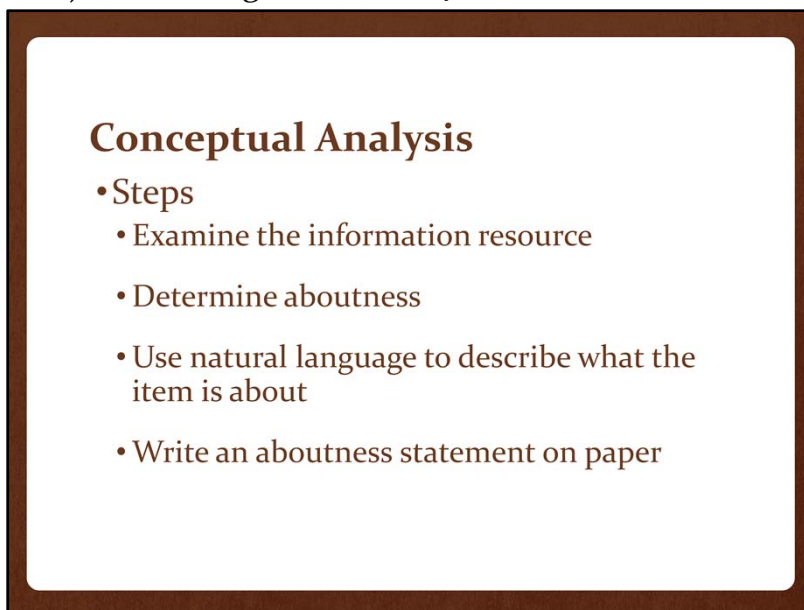
Conceptual Analysis
- Steps
  - Examine the information resource
  - Determine aboutness
  - Use natural language to describe what the item is about

No matter what methods a cataloger chooses to use, determining the aboutness begins in the natural language. We perform analysis independent of any subject languages, that is, independent of a controlled vocabulary, or a classification scheme.

Now, why is this important? Well, *all* controlled vocabularies have context. They are designed to collocate what the scheme's creators think should be brought together. So they reflect the culture and the point of view of those creators. For example,

- Should a particular class of persons be called rebels? Or are they freedom-fighters? Or terrorists?

- Do we use the term pro-life or anti-abortion? Pro-choice or pro-abortion?

- Within classification do Art and Music belong together? Or should we put Music and Math together, instead?

- In the classification system, how much room should be left for Christianity versus other religions?

Creators of systems make choices, and there are always alternatives to the choices that are made. We do not want to consider systems when we consider conceptual analysis. If we approach subject analysis with a particular system in mind, it can be very limiting. But, if you've been using a system for a long while, you *really* know the system well, and you can't help but be influenced.

Once you have an idea of the subjects, the genres and forms, and other characteristics of the resource, then it is important to sort out your thoughts in the form of a statement.

Creating an aboutness statement allows you to take into account the most important concepts to include in your translation.

You want to approach it with a simple sentence that begins *"This resource is about…"* or *"This book or video or website or whatever is about..."*

Writing it down is important; not doing so can allow some details to be lost.

Let's examine an actual resource and see how this could be done.

The title is *Introduction to Cataloging and Classification*.

It is a textbook that is 1,048 pages long, that was written in 2015 by two cataloging professors and a librarian who manages a technical services department.

It has six parts, comprising twenty-four chapters.

Examining a work

Now let's look at the table of contents.

We can see that there is an introduction as Part One.

Part Two, "Description and Access", begins on page fifty-one and runs all the way at least through page 465. That is four hundred pages; that is a significant portion of a thousand-page book.
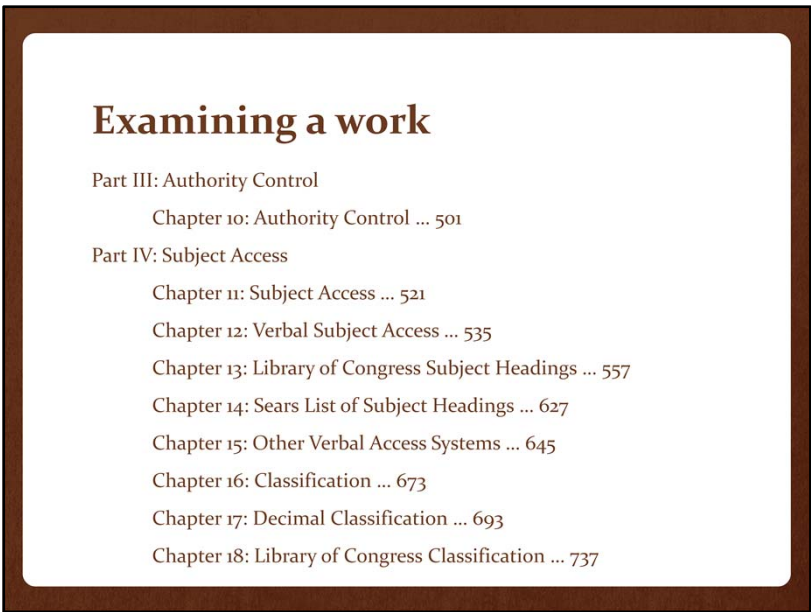
In addition, we can see that RDA is covered. We have a chapter on RDA Basics, and then we're looking at this and seeing Works and Expressions, Manifestations and Items, and so forth. All of this looks as though it relates to RDA. We'll keep that in mind.

Examining a work
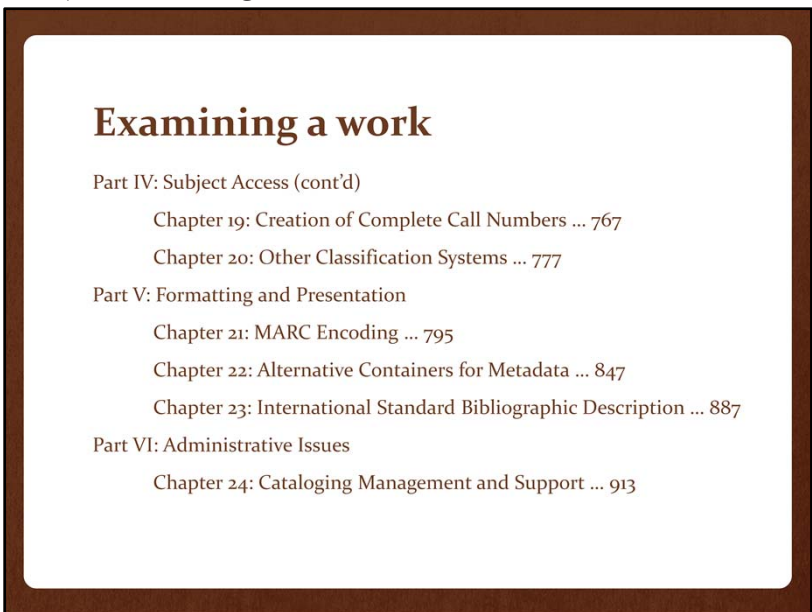
Continuing to look through the table of contents, we see there is a relatively short portion on authority control. It's only about twenty pages.

But then, Part Four for "Subject Access" is at least two hundred pages, and it includes several references to subject headings, particularly LCSH and Sears (which is another system that is often used in public libraries and school libraries) and we also see classification mentioned several times.

Examining a work

There is one more page to the table of contents. We can see that "Subject Access" has been continued onto the new page, with chapters on the creation of complete call numbers and of other classification systems.
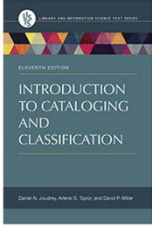
Part Five is "Formatting and Presentation". This involves the way that we encode our bibliographic and authority data. And you can see it's about 120 pages, which is *significant*, but the question will be, is it significant enough to be included in the subject analysis.

There is also a relatively short section on "Administrative Issues".

So we can see that this appears to be a textbook on all of the procedures for cataloging, with an emphasis on descriptive and subject cataloging and a special notice given to RDA.

Aboutness Statement Example

This book is about library cataloging. It includes a significant section on descriptive cataloging, including an extensive explanation of the various parts of *RDA: Resource Description & Access*. There is another large section on subject cataloging, focusing on subject headings work and classification.

This is the aboutness statement that I came up with:

This book is about library cataloging. It includes a significant section on descriptive cataloging, including an extensive explanation of the various parts of *RDA: Resource Description and Access*. There is another large section on subject cataloging, focusing on subject headings work and classification.

Subject Analysis Process

• Conceptual analysis of item

• Translation of the aboutness into specific notation or terms

Now that we have a good idea of what the resource is about, we have to determine what topics are most important.
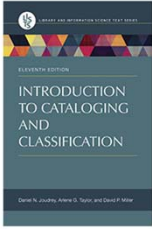
We must think about

• prioritizing the concepts,
• familiarizing ourselves with the subject heading list,
• sketching out a classification hierarchy, and
• picking words to search in the vocabulary.

We will now need to translate that aboutness into one or more subject languages, like LCSH and LCC.

Aboutness Statement Example

This book is about library <u>cataloging</u>. It includes a significant section on <u>descriptive cataloging</u>, including an extensive explanation of the various parts of *RDA: Resource Description & Access*. There is another large section on <u>subject cataloging</u>, focusing on <u>subject headings</u> work and <u>classification</u>.

We will now go back to our statement and mark the significant concepts.

We will then translate those concepts into controlled vocabulary.

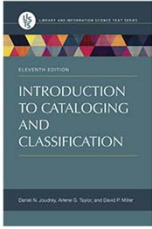On the slide, various terms are highlighted, and those are the terms that we will search in LCSH.

These controlled headings represent the aboutness of the textbook:

- **Descriptive cataloging**,
- **Subject cataloging**,
- **Classification—Books**, and
- **Resource description & access**.

Now you might ask, "Why isn't the heading **Cataloging** there? That's the whole point of the book." In this case the cataloger decided to give narrower headings for each function of cataloging instead of a more general heading, in an effort to provide more specific access.

Future training modules will provide you with practice in making these decisions for yourself.