To Dream the Impossible Dream, To Read the Unreadable Script: Cataloging Non-English Books

By Richard Allen Lammert

How many non-English books do you have in your backlog? More than two? How many are in the backlog because you do not know exactly how to approach working with them? How many are in a non-Roman script? Many libraries do not have a dedicated employee to catalog non-English books. So when it comes to cataloging those materials, there are three basic approaches: push the book back into the backlog, get someone else to catalog it, or learn to do it yourself. My approach has always leaned toward "do-it-yourself."

My own interest in cataloging non-English books began as a personal interest, but it became very much an institutional necessity when I began working at Concordia Theological Seminary, Fort Wayne, Indiana, in 1998. The Lutheran Heritage Foundation of Macomb, Michigan, sends us free copies of the books they translate, publish, and distribute, in order to represent them in a library. These books mostly relate to Lutheranism and fit our collection development policy as primary material that we would normally acquire. Some of the books are original publications, but many are translations. There are currently publications in 105 languages, in a number of different scripts. Since very few libraries are obtaining these publications, the only sure way to get them into WorldCat—and our own catalog—is to "do it myself." Although the foreign-language material from the Lutheran Heritage Foundation is not the only such material we obtain, it is the bulk of such cataloging.

Because I thought others could benefit from what I had learned, I offered preconference workshops at the 2003 and 2011 Atla annual conferences called *Tackling the Foreign-Language Backlog*. I always thought it would be useful to the profession to take the notes that I had created for these workshops and expand them into a book. The Atla Press has agreed to publish this book with the same title as the workshops.

The subtitle, *Transcription and Romanization for the Accidental Foreign-Language Cataloger* indicates that the book covers only descriptive cataloging, not subject cataloging, of non-English books. The questions about cataloging foreign language materials that appear on electronic discussion lists are usually of the type that I cover in this book: "Can you help me identify the script of this book?" "Can you provide me a romanization of this title?" My presupposition is that you are aware of the subject matter in a book that comes into your library, so I don't try to provide detail on possibilities for translation (although they do get mentioned). It also indicates that it is geared for catalogers for whom foreign-language cataloging is only an occasional part of their workload. (And if you *do* know a particular script, don't use my book for that script—not only is it unnecessary for you, but it will only confuse you.)

The first section of the book covers cataloging in Roman scripts. Although easier than non-Roman-script cataloging, one can encounter difficulties here. Two examples will suffice. The first example is with the Slovakian word veľké? If you are familiar with the MARC diacritics, you may think that the third letter is a lowercase L with high comma, off center. A check of the chart of modified Roman letters I have included in the book shows that this is actually a lowercase L with hacek (˘) above it. Slovak (as well as Czech) typography tries to avoid putting a hacek above the lowercase L since the hacek pushes into the line above it.

Richard Lammert is Technical Services Librarian at Concordia Theological Seminary.

Connexion Client will allow you to enter a lowercase L with high comma, off center, creating a character that may look similar (depending on the font selected) as a lowercase L with a hacek. Entering a lowercase L with hacek, on the other hand, is the correct diacritic to use, regardless of the font selected.

	HIGH COMMA	**HACEK**
TIMES NEW ROMAN	ľ	ľ
ARIEL UNICODE	ľ	ľ

You may think that clearly you need to use the first method (with high comma). Despite the way the character is displayed in Connexion Client, Slovak online catalogs input this word with a hacek, and a good browser will *display* it as a lowercase L with high comma, off center. Currently, most catalogs (including that of the Slovak National Library) will retrieve the word with or without a diacritic (even a wrong one). But as libraries enter the new global world of linked data, I believe it makes sense to enter our data so that it conforms to standards outside the Anglo-American cataloging community.

A second Roman-script example is working with Fraktur script. There are certainly web sites that help you with this, but I put the charts and material right in the book for you, with some focus on the understanding of Fraktur script in a bibliographic environment. This, like much in the book, can be found elsewhere if you know where to look, but I think that putting it all together in one place will benefit catalogers.

Most of the book is on romanization or transliteration of non-Roman (Latin) scripts. The beginning chapter of this section provides an introduction to romanization, covering the philosophy and practice, as well as the problems and concerns. Although romanization was used to provide a way to file printed catalog cards in a Roman-script catalog, romanization has been continued into the computer age, where it is, strictly speaking, not necessary. Nevertheless, we still use romanization, so we have to work with it. This introductory chapter is followed by a presentation of twenty-nine scripts from around the globe. Similar scripts are grouped together with sections on scripts of Europe, scripts of India and South Asia, scripts of Southeast Asia, scripts of East Asia, scripts of the Middle East, and miscellaneous scripts.

The ALA-LC romanization tables are a necessary starting place for doing any romanization work, but they are not written for someone who is not familiar with the language and script. For instance, if you check the romanization table for Hindi, the first thing you read after seeing all the consonants and vowels spelled out is this sentence: "Only the vowel forms that appear at the beginning of a syllable are listed; the forms used for vowels following a consonant can be found in grammars; no distinction between the two is made in transliteration." Now, if you already knew Hindi, this would be no problem. But, for an accidental cataloger of Hindi, this is exceedingly unhelpful.

Fortunately, listing out *all* forms of consonants and vowels is not particularly onerous. You will find these explicitly listed in the book and not have to consult a grammar for them. Each ALA-LC romanization table (not just the one for Hindi) assumes a certain knowledge of the language and script. In *Tackling the Foreign-Language Backlog*, I try to spell out all the assumptions, and make everything explicit.

The inability to find what is needed in the ALA-LC romanization tables is not just a theoretical assumption on my part. A few months ago, I saw a posting on a cataloging list from a cataloger who was trying to romanize the title of a web site in Tamil. He noted in his posting, "I tried using the ALA-LC Romanization table, but it does not include the characters I need." The characters were not included because the Tamil table had the same note as the table for Hindi, that the forms of vowels used with consonants were not listed.

Another challenge when trying to use the ALA-LC romanization tables occurred when I was trying to use them for romanizing Sinhala. I was already aware of the various forms of the vowels in this language, so that wasn't a problem. However, every time I transliterated a title I was working with, I ended up with a *different* romanization. I knew that was not possible.

The problem, as I finally worked it out, was that I was confusing different letters that looked very similar. Those letters were in different places in the romanization table, because, in the alphabetic order of Sinhala, they were separated by a number of other letters. Sometimes I would find the match with the first letter and have one romanization; at other times I would find the match with the later letter, and have a different romanization. I realized that the solution was to reorder the letters so that similar letters were *together* in my table. Since the accidental foreign-language cataloger does not know the alphabetic order of the letters, this makes no difference. Yet it helps one not familiar with those letters to find a proper match, while at the same time avoiding reading a letter as a different, very similar letter.

In order to provide actual practice in romanization, I include an example of a title in each script, demonstrating the process of romanization. I point out common mistakes that can be made in reading a particular script so that you can avoid making the same mistakes. Most of the mistakes I illustrate are actual mistakes I have made at some time or other, so they are authentic.

Although I have not worked in *every* script I include with actual books in my own library, I have used most of the materials in the book. The book is thus user-tested—even if it was by only one user. Thus, I have become my own client, using the information and charts to do my own work of cataloging. I know that what I have included in the book is usable and works, since I use it myself.

In addition to the free PDF copy of the book, a print copy of the book will also be available. I actually designed the book to be used in print format, since many of the tables occupy two facing pages, and aren't complete without both pages. You can certainly view these pages in PDF format, but a printed copy (or a printout of the pages) puts the full table on both pages in front of you, without the necessity of moving between the PDF pages.

I would certainly be interested in hearing of your use of the book, especially any problems you encounter with its use, or situations that I have not anticipated. I would like to improve the book in any way possible. The book should be available later this year (hopefully) and an announcement will be made when it is published.