OLAC CAPC
Video Language Coding Best Practices Task Force
Draft Recommendations

Members:
Kelley McGrath, chair
Cindy Badilla-Melendez
Susan Leister
Katia Strieck
Carolyn Walden

Introduction

The task force was charged with creating a set of best practices for coding MARC 008/lang and 041 language information for videos, especially DVDs, and with using that exercise to examine whether any changes could be made to the MARC format (coding or directions) that would improve access to the multiple types of language information found on videos.

The task force's recommendations are based on the following premises:

1. Coded language data is intended for use in retrieval, limiting, and sorting.

2. Coded language data does not need to describe all language-related information about an item that might be of interest to users. Coded language information can be expanded on and complemented by information in 546 free text language notes.

3. Coded language data is most effective when it supports the retrieval of the language(s) of the main work(s) on the item, rather than the language(s) of supplementary and bonus materials.

4. Coded language data is most effective when it supports retrieval based on language(s) in which the item is usable, rather than all language(s) that might be found in the item.

5. For moving image materials, patrons are most interested in the retrieving, limiting, and sorting by the following types of language information:

   - spoken language of the main content
   - written language of the main content (including subtitles, captions, and intertitles)
   - the original language of the work

After examining the types and combinations of language information that occur in videos, we created a list of types of language information that we thought were important for retrieval and a list of those that we think can be adequately expressed in free text notes in 546. We then focused our efforts on how best to make spoken, written, and original language information consistently and effectively accessible.

Below we have listed some recommendations for general practice, particularly for what we perceived as tricky situations, and for possible changes to the MARC format to support better use of coded language data for moving images. Note that the examples use the not-yet-implemented

041 subfield j, which was approved in May 2007 for use for subtitles and captions instead of including them in subfield b with summaries and abstracts (see http://www.loc.gov/marc/marbi/2007/2007-01.html)

Types of Information Not to Include in 008/lang and 041

- Packaging language(s) (disc or tape label, container, disc menu)
- Special feature language information (audio commentary tracks on DVDs, spoken and written languages on special features)
- Credits
- Accompanying material (e.g. guides, booklets)

This last needs a bit more discussion, as currently the language of accompanying material can be explicitly coded in 041 $g (defined as language code of accompanying material other than librettos, or scripts or accompanying sound for visual materials) when the material is considered significant. It is not clear to the task force when this would be useful for retrieval. Unlike musical recordings, few videos have significant accompanying material and, in practice, this subfield generally does not seem to be recorded if both the accompanying material and the video are in English. It seems unlikely that most users would want to search separately for the videos with accompanying material in a certain language (although it might be useful contextual information once they are looking at an individual record) nor that we would have enough data in such an index to support a useful search option. We are also concerned about the use of $h (language code of original and/or intermediate translations of text) after $g, although it appears to be rarely used for moving image materials, because it decreases the ability to use $h to help determine the original language of the main material.

We therefore recommend that the language of accompanying material be mentioned in a note if deemed important, but do not think it is necessary to record it in coded form in 041. However, there seem to be no negative impacts from coding $g if it is so desired. Therefore, we recommend that $g be used for accompanying material at the discretion of the cataloging agency, when the cataloger deems that it is important and useful to code, but that it not be required.

Do not use $h following $g, even if the accompanying material is a translation and the original language is known.

Examples of Coding Ignoring Packaging and Credits

Swedish film that has been dubbed into English; credits (except for title) still in Swedish. Packaging and menus in English

> 008/lang eng
> 041 1# $a eng $h swe
> 546 ## Dubbed in English; credits in Swedish.

Our Daily Bread. No spoken dialogue, no intertitles, no subtitles; English credits, menus, and packaging; originally produced in Germany.

> 00/8lang zxx
> 546 ## English language-credits version of a German film without dialogue.

Il Cerchio = The Circle. Edizione italiana. Farsi soundtrack (original Farsi); English subtitles; Italian credits on screen.

> 008/lang per
> 041 1# $a per $j eng $h per
> 546 ## Farsi soundtrack with English subtitles. Onscreen credits in Italian.

Spoken and Written Language Information

If we would like to be able to search spoken and written languages separately, the current definition of 008/lang and 041 $a for moving images creates ambiguous data. The definition seems to be intended to code the "main" language of the item, reverting to written language if there is no spoken language. We recommend coding 008/lang and 041$a only for spoken languages and using the MARC code zxx (no linguistic content) for videos with no spoken language. We recommend including intertitles in the definition of $j with subtitles and captions. We also recommend that sign languages be included with other non-spoken language information in $j. This would enable the separate retrieval of spoken and written languages when desired, as well as the creation of an index for "accessible in" languages that would include both 008/lang and 041$a plus 041$j.

Standard Examples

Original English dialogue; English packaging, menus, and credits

> 008/lang eng

Original English dialogue; closed-captioned in English; English packaging, menus, and credits

> 008/lang eng
> 041 0# $a eng $j eng
> 546 ## Closed-captioned.

Japanese language film; English subtitles; English packaging and menus

> 008/lang jpn
> 041 1# $a jpn $j eng $h jpn
> 546 ## In Japanese with English subtitles.

Japanese language film; optional dubbed English soundtrack; optional English subtitles; English packaging and menus

> 008/lang jpn
> 041 1# $a jpn $a eng $j eng $h jpn
> 546 ## Japanese or dubbed English soundtrack; optional English subtitles.

English language film with English, French, or Spanish soundtracks; closed-captioned in English; optional subtitles in English, French, Spanish, Portuguese, Chinese, or Thai. English packaging and menus

> 008/lang eng
> 041 1# $a eng $a fre $a spa $j eng $j chi $f fre $j por $j spa $j tha $h eng
> 546 ## Closed-captioned; English or dubbed French or Spanish soundtrack; optional English, French, Spanish, Portuguese, Chinese, or Thai subtitles.

Recording of *The Bridge*, an opera performed in sign language, simultaneously sung in English

> 008/lang eng
> 041 0# $a eng $j sgn
> 546 ## Performed with gestures, American Sign Language, a musical soundtrack, and in English.

Examples of Films with No Spoken Content

Symphony performance; no spoken/sung language; credits in German; disc menu and packaging in English.

> 008/lang zxx

A Chaplin silent film on DVD with multiple subtitle tracks

> 008/lang zxx
> 041 1# $a zxx $j eng $j chi $j fre $j kor $j por $j spa $j tha
> 546 ## Silent film with English intertitles and musical acc.; optional French, Spanish, Portuguese, Chinese, Thai, or Korean subtitles. Optional audio commentary track in English. Menus in English, Spanish or Portuguese.

Original language

Options

The task force feels strongly that it is important to provide access to the original language of moving images. Users in many situations are interested in films that were originally in French, Spanish, Arabic, etc. and we do not currently have an effective, standard way to provide this information. The task force came up with three possible methods for addressing this:

1. Create a new subfield in MARC 041 for the original language(s) of the video or film (or other work).

   We do not think the existing subfield $h can be made to work reliably for this purpose because a) it is defined to include languages of "intermediate translations" and b) since it is only used when something is a translation, if there is no $h, it is necessary to try to infer the original language from 008/lang and 041$a. However, for this to work, it is necessary to assume that catalogers always add an 041$h when it applies, which in practice is not true and thus unlikely to produce sufficiently reliable results.

A new subfield would be unambiguous and would only have to be used by those who want to use it. In the case of film and video, it seems unlikely that there would be many languages of "intermediate translations" so OCLC could probably populate this new subfield with the content of existing 041$h languages for moving images.

If this were defined as $o for original language, it might look like:

041 0 $a eng $o eng

041 1 $a spa $a eng $a fre $b eng $b spa $h spa $o spa [although the $h does become somewhat redundant with this practice]

2. Lobby the Library of Congress to implement LCSH genre-form headings that include original language, e.g., something like "Feature films $x Spanish language" or "Feature films $x French language $z Cameroon." "Foreign language films" is too relative to be useful here.

3. Create work-level records that include original language and other work-related information.

By this we mean the development of some sort of record for the work that includes more information than current authority records. In the long run, this might be the most effective and efficient way to add this sort of information in a shared environment, but it would take longer to implement as it would necessary to set up a new infrastructure. This could be developed as something with lower entry barriers than NACO, but with some sort of provisions for quality control, such as saying where each bit of data came from so that the records could be improved over time and it would be easy to identify places where a cataloger has guessed at something or taken information from packaging so that the information could be confirmed in more authoritative sources. These work records could be linked to corresponding ISBNs, OCLC numbers, etc. and used to enhance local catalog access or to create new interfaces (say something like OCLC's Fiction Finder, http://fictionfinder.oclc.org/)

Data elements could include:
Title: Casablanca
Language: English
Date: 1942
Director: Curtiz, Michael, 1888-1962.
Country: United States
[could be expanded with other information such as cast, other production personnel, summary, subject and genre headings, setting (place and time period), awards, color, sound, aspect ratio, other technical details]

Recommendation

Although LC is currently working to implement LCSH genre-form headings for moving images, they are not planning to make any provision for the inclusion of language. Therefore, we recommend proposing a new 041 subfield for original language, as this is the simplest and quickest approach to implement. We will also ask CAPC to examine the issues related to developing work-level records for moving images.

Other Issues

Multiple Works with Different Language Information on One Bibliographic Record

Since 041 is a repeatable field, use separate 041 fields when needed for different works on one manifestation. Optionally, based on cataloger's judgment, if the language descriptions for the works represented in one manifestation are numerous, code tag 041 for as many works as practical and use the code mul (multiple languages) when necessary.

Shorts! Volume One : 15 Award-Winning Film Festival Shorts

> 008/lang eng
> 041 0# $a eng $j eng
> 041 1# $a dut $j eng $h dut
> 041 1# $a dan $j eng $h dan
> 546 ## Selected movies are closed-captioned; soundtracks principally in English; The Chinese wall in Dutch with optional English subtitles; John and Mia in Danish with optional English subtitles.

Disc 2 of *The Wild West*. Includes the Italian film C'*è Sartana* (*Fistful of Lead*), dubbed in English, and the English language television feature *The Gunfighters*.

> 008/lang eng
> 041 0# $a eng [not normally necessary]
> 041 1# $a eng $h ita
> 546 ## Fistful of lead in dubbed in English. The gunfighters in English.

Videos with Brief Sequences in Language(s) Other Than the Main Language

Include brief or subsidiary languages in 546 if thought important. Code lang/041 for languages which are substantial and which the intended audience needs to be able to understand to use the item.

The Internet Movie Database tends to list all languages (see their record for *The Godfather* at http://www.imdb.com/title/tt0068646/), even from brief sequences; DVD packages tend to list soundtracks by the language(s) of the intended listener (except in the case of originally multi-lingual films described in the next section). We believe the latter approach is more useful for most users of library catalogs.

The Godfather is primarily in English, but has a few Italian sequences and apparently some in Latin.

> 008/lang eng
> 546 ## In English with brief sequences in Italian and Latin with English subtitles.
> (or no 546 if not deemed necessary to note Italian and Latin sequences)
> no 041 (rather than 041 0# $a eng $a ita $a lat $j eng; ignore the Italian and Latin for coding purposes)

Videos with Mixed-Language Soundtracks

When substantial portions of a video are in more than one language, code for all substantial languages present.

An Algerian DVD that is a clear mixture of French and Arabic. The characters often switch between the two languages within a sentence and depending on who they are talking to use either French or Arabic. No subtitles.

> 008/lang ara (since the package and credits were in Arabic)
> 041 0# $a ara $a fre
> 546 ## Dialogue consists of a mixture of Arabic and French.

This means that sometimes users searching by spoken language will retrieve videos that are not usable to monolingual speakers of languages coded in 008/lang and 041 $a. In the examples of *Joyeaux Noel* and a hypothetical DVD below, both have the same information in 008/lang and 041 $a, but only the second example would be useful to a monolingual speaker of French or German. The task force sees no way to compensate for this problem in the existing MARC language coding structure nor do we believe that there are existing user interfaces that are sophisticated enough to deal with these distinctions.

Joyeaux Noel. Soundtrack of DVD and original film in English, French, and German; optional English, Spanish, or Portuguese subtitles.

> 008/lang fre
> 041 1# $a fre $a eng $a ger $j eng $j por $j spa $h fre $h eng $h ger
> 546 ## Soundtrack in a mixture of French, English and German; optional English,
> Spanish, or Portuguese subtitles.

A hypothetical DVD of a French film that has optional English, French, or German soundtracks and English, Spanish, or Portuguese subtitles.

> 008/lang fre
> 041 1# $a fre $a eng $a ger $j eng $j por $j spa $h fre
> 546 ## French, English, or German soundtracks; optional English, Spanish, or Portuguese
> subtitles.

Recommended Changes to the MARC Format

1. Change the instructions to limit 008/lang and 041$a to spoken languages for moving image materials. Use MARC language code zxx for items with no spoken language content.

2. Change the instructions to include intertitles on silent films and sign language on videos in 041$j along with subtitles and captions

3. Create a new subfield to unambiguously code the original language of the main work(s) on a record

Appendix: Captions, Intertitles, and Subtitles

Functional definitions of terms used in this document

Intertitles: Generally associated with silent films, intertitles usually appear as a separate frames containing written dialogue or other information to aid in comprehension.

Subtitles: Subtitles are text, usually appearing at the bottom of the screen, that provides a translation or transcription of spoken dialogue. Intended for viewers who can hear the soundtrack, subtitles are usually used for translations of foreign language films.

Captions: Captions are similar to subtitles, but also include contextual clues for viewers who cannot hear the soundtrack, such as identification of the speaker when it's not clear from the action on screen, and sounds, such as explosions or phones ringing.

Function vs. Encoding Method for Subtitles and Captions on VHS and DVD

For VHS videos there is generally a 1:1 correspondence between the function and the encoding format of captions and subtitles.

VHS

	Function	Encoding format
Closed-captions	Caption	Embedded in video signal. Requires hardware (line-23 decoder in VCR and in TV; included in all U.S. TVs 13" and over manufactured since July 1993)
Open-captions	Caption	Imprinted on the tape; usually look blocky like closed-captions, but cannot be turned off and do not disappear when tape is fast-forwarded
Subtitles	Subtitle	Imprinted on the tape; cannot be turned off

Unfortunately, for DVDs this is not the case and there is sometimes no practical way to tell if a DVD has subtitle or captions.

DVD

	Function	Encoding format
Closed-captions	Caption	Embedded in video signal. Requires line-23 decoder in DVD player or drive and in TV (included in all U.S. TVs 13" and over manufactured since July 1993). Not all DVD players or drives, especially older models, include the necessary decoder. In addition, the way the DVD has been encoded can interact with particular hardware and software configurations, such that the line-23 captions do not work (some captions will play on a stand-alone DVD player with TV, but not on a computer DVD drive or vice versa)
Open-captions	Caption	Cannot be turned off; may be encoded as part of the film image
Subtitles	Subtitle	Cannot be turned off; may be encoded as part of the film image
Optional subtitles	Subtitle	Digital subpicture bitmap overlay (not possible on VHS; usually turned on or off from the disc menu or remote, but sometimes hard-coded by the DVD producer so that they cannot be turned off)
Optional subtitles for the deaf and hard of hearing/captions	Caption (Sometimes referred to by publishers as SDH, "subtitles for the deaf and hard of hearing," or "English captions." Sometimes called "English subtitles" even though captioning information is also included)	

Resources For Further Information on Subtitles and Captions on DVDs

http://www.dvdfile.com/site/faq/caption_guide/
http://www.dvddemystified.com/dvdfaq.html#1.45
http://joeclark.org/access/dvd/capabilities.html
http://en.wikipedia.org/wiki/Closed_captioning
http://en.wikipedia.org/wiki/Intertitle
http://en.wikipedia.org/wiki/Subtitle_%28captioning%29