

May 18, 2003

Kasparova, N.N.

Bibliographic Record Language in Multilingual Electronic Communication: to International Cataloguing Rules Conception

The bibliographic record is the main element of an information bibliographic system, and in the electronic environment the language of a bibliographic record facilitates communication between a user request and the contents of bibliographic system. The text of the bibliographic record in natural language allows online searching of bibliographic records by the names of persons, by titles, by subjects, or by keywords found in the bibliographic records. Therefore the language of the bibliographic record together with its script are important elements in connecting the users with desired documents” and should be controlled by special rules to enable good search and retrieval. The duality of bibliographic information, for both description and access, becomes obvious on the level of its main element – bibliographic record. When creating rules for choosing the language of the bibliographic record it is necessary to define as precisely as possible the logical scheme of choosing the language in order to avoid a discrepancy between two tasks: reflecting accurate description of the item on the one hand and maintaining controlled access to it for local and foreign users on the other.

So the existence of multilingual catalogues, especially catalogs available through the world-wide Internet, reflects the multicultural nature of the universal bibliographic information as a whole and yet results in a conflict between a task to provide searching for local users and a task to give accurate information about a document that is understandable to foreign users. The language and script of the bibliographic record have multiple requirements to connect multilingual users with documents.

The active development of ISO international standards for transliteration shows the intention to solve a part of this problem by romanization of different non-roman scripts thereby eliminating the variety of scripts that a user will need to deal with. But a such model may exist as a parallel or provide additional access, for users who are accustomed to Latin script languages, while allowing the main access points and bibliographic description to be in the language and script of a the document for all people in the world.

Each cataloguing code contains rules concerning the language of a bibliographic record. Those rules usually reflect a focus on the needs of the catalog’s local users. Mostly the cataloguing rules are oriented on a local user, sometimes against foreigners, to facilitate access to multilingual records for a library’s own users or, in the case of some national libraries, for their own citizens.

The best known foreign cataloguing rules in Russia – Anglo-American Cataloguing Rules (AACR) – demonstrate the evident preference of English language for notes and qualifiers to headings, to all others especially “exotic,” or non-roman script languages, which include together with Chinese, Arabic, etc. all Cyrillic languages including Russian. The romanization of non-roman scripts is used to facilitate the data retrieval for English-speaking users.

This problem is decided more prudently in the current Russian (and former USSR) cataloguing rules where using of transliteration or transcription is recommended rarely. But these recommendations are not clear, nor motivated and oriented on concrete goals.

The problem of choosing of a bibliographic record language is not solved also in MARC formats. There are no instructions on using languages and scripts for recording data in each field or subfield, but the format gives some possibilities to record data in different languages (fields for parallel titles and translated titles) and different scripts (special field 880 in MARC 21). This is further argument to confirm that cataloguing rules are systematic means of maintaining the intellectual communication between a document and a user while a

May 18, 2003

format supports formal communication between a user and a computer. Electronic communication, on-line catalogues and technology of creating of machine-readable records, give now new possibilities to solve the problem of choosing the bibliographic record language and script more harmoniously and impartially as demonstrated by the following:

1. The powerful and dynamically developing international standard, UNICODE, where almost all scripts are represented, allows entering bibliographic data in the script of a title page. And this is more objective reflection of a document and corresponds to international interests and to international bibliographic description rules, like ISBD. It is a pity that not all automated bibliographic systems yet support UNICODE, but this is a matter of time: more and more vendors are on the way to developing their systems with an understanding that the progress of the world as a multicultural and multilingual society is irreversible.

2. The technological possibility is closer to a reality of presentation and transmission of electronic bibliographic data based on the MARC format and linked to authority files of different multilingual data (personal names, corporate names, geographical names, uniform titles), and all these together with convenient software. The structure and content designation in the MARC format allow representing data about a document both in its own language and script as well as in another, and together with multilingual authority files searching possibilities increase significantly in bibliographic systems. One problem to reaching this goal is the poor development of authority data systems in some countries including Russia. Moreover, even existing authority data are not always accessible for many libraries due to costs, and sometimes because their library's own software doesn't support authority control.

To more logically present possible international rules on choosing language and script for bibliographic records, it is necessary to take into account the role of each bibliographic record element as it will be used in electronic communication.

Let us examine these points in details:

1. Title and responsibility statement.

Title is the main access point for a work, and that's why it should have the highest priority as an access point for international users. Transcribing the title as it is found in the item allows for an objective reflection of the item itself. It follows that the language of the title statement in a bibliographic record should objectively display data about a document and correspond to the language and script of its imprint. But if it is decided to give more title access points for local and international use, the MARC formats provide this possibility as additional access points. The most used international formats, MARC 21 and UNIMARC, include special fields for data in another script and for a translation of title by a cataloguing agency. The Russian Cataloguing Rules follow this practice, and it is suggested for the international cataloguing rules created under IFLA's Cataloguing Section. (ISBD already allows this and most cataloging codes do, too.)

2. Imprint statement.

Following the same logic for the edition statement, imprint, and series statement, all of these data should be given in the language and script of the title page. (ISBD and most cataloging codes do this, too.)

Аналогичный подход к сведениям об издании, выходным данным, и области серии. Все эти данные приводятся на том языке и графике как они приведены на титульном листе документа.

3. Notes.

The language of notes formulated by a cataloguer should correspond to the language of cataloguing or to the language of the cataloguing agency, which is the language of a country. (Some countries have two or three or more official languages – Canada 2, Finland 2, Switzerland 3, UNESCO has 10!, etc.) These data are intended for local users, for example, Russian language in Russia, German in Germany, etc. It is difficult to imagine that a cataloguer will formulate such notes as numbering or typographic peculiarities in the language of a document, sometimes little known both for a cataloguer and for a home user. Notes mainly contain data helping to give additional information about a document and about ~~an~~ access to it. That's why they should be directed to the language needs and expectations of a local user, with the exception of an original version note where data have been given in the language and script of the title of the object of cataloguing. (ISBD and most cataloguing codes already cover this, too.)

4. Subject access language.

Subject headings, keywords, classification numbers provide access points to search for documents on a subject or topic of interest. Excluding numeric or alphanumeric classification numbers, subject headings and keywords are represented in a natural language in a record. Accordingly a user may formulate his/her subject request in a natural language, usually his/her native language. In a bibliographic record subject access data are taken from local subject headings and standard thesauri or from similar translated materials. These data are intended to be for local users and should be represented in their language independent of the language of the document. If there is a possibility to translate and to maintain these data in two or more languages or there is an authority control system through multilingual authority files of subject headings, search possibilities of a bibliographic system will be enriched considerably. Subject access becomes more harmonized, friendly for both local and foreign users. (This reflects current practices in most countries.)

5. Bibliographic headings.

The choice of language for bibliographic headings is more complicated than the simple transcription using the language of imprint data. Qualifying information, like titles of nobility, terms of address, academic degrees, etc. may originate from sources other than a described document and are represented in the language of cataloguing – the local user's language. Multilingual authority files of names, uniform titles, geographical names also help with offering display choices. But sometimes a country or a library doesn't have access to authority data and should make a choice based on the type of heading and its function in an electronic environment. In modern technology of machine-readable cataloguing, a heading loses his former role of a card catalogue organizer because this problem may be decided by other tools in the electronic environment. **(Still need to use for ordering displays in the online systems???) Being an optional element (???)** a heading is one of the access points formulated by a cataloguer and may be different from the data of a document in its structure, form, language, and script. I would like to underline once more that heading form and structure are controlled through corresponding authority files and by special rules (standards) but the choice of language for these data may be diverse (that is more free) at a bibliographic agency's discretion, according to its user service needs.

Более сложно и неоднозначно решается проблема выбора языка заголовка библиографической записи, хотя общий подход соответствия языку выходных данных документа сохраняется, за исключением идентифицирующих признаков (указание духовного сана, и т.п.), которые формируются не с документа, а, как правило, из других

источников и соответствуют языку каталогизации – языку национального пользователя. Проблему выбора языка этих данных помогают решить многоязычные нормативные /авторитетные файлы имен, наименований учреждений, анонимных классических произведений, унифицированных заглавий, географических наименований. Однако, не всегда страна или библиотека имеет доступ к нормативному файлу, в связи с чем вынуждена делать более сложный выбор, учитывая статус заголовка библиографической записи и его функции в электронной среде. В современной технологии машиночитаемой каталогизации заголовки утрачивают свою былую функцию организации массива в карточном каталоге, поскольку в электронной среде эти задачи решаются другими средствами. Будучи факультативным элементом, заголовок библиографической записи является одной из точек доступа, которая формируется каталогизатором и может отличаться от структуры, формы, языка и графики данных, приведенных на документе. Еще раз подчеркну, что форма и структура заголовка контролируется соответствующими нормативными/авторитетными файлами и специальными правилами (стандартами), однако, выбор языка этих данных может быть различным, то есть более свободным, на усмотрение библиографирующего учреждения в соответствии с задачами обслуживания или структуры справочного аппарата данного учреждения.

5.1. Name of Persons in Heading

Cataloguing rules of different countries give many types of motivation for choosing a language of personal name in a bibliographic record heading. For example, a language of:

- the majority of an author's publications,
- the first appearing publication,
- fatherland of an author,
- a document.

Different methods of transliteration are also suggested. This variety of data presentation is a cause of many forms of a personal name appearing in bibliographic records, and a user searching by author's name in different catalogues would need to write them one way or another. For example, today a Russian user would need formulate his/her request in romanized form if he/she searched by that Russian author's name in the Library of Congress catalogue, and perhaps h/she would be forced to test several versions of romanization of this name, if the one chosen by the Library of Congress was not known to the user. The best way to resolve this problem is in creating multilingual authority files. In this case a user knowing an author's name from the title page form has a possibility to find it in such a file and then executes next steps of searching. So it is necessary to come to an agreement to give author's name in the language and script of a document in the responsibility statement (ISBD area 1) and in the corresponding authority file record. Then a choice of language and script of an author's name for a bibliographic record heading may be left to a bibliographic agency, but the local user's language and the author's native language would be given a higher priority. Transliterated forms of a name may be given as it is recommended by Anglo-American Cataloguing Rules for Cyrillic, Chinese, Japanese, etc. But the obligatory condition is observing international standards for transliteration. For example, bibliographic record headings on publications of translations of Russian poet Pushkin, A. S. will be given in Cyrillic in Russian catalogues as a rule, but it may be desirable to give his name in other scripts in an authority record for his name. Anyway this name in the language of publication will be given in the responsibility statement and this is obvious. The same is for corporate name headings.

Правила каталогизации различных стран дают множество мотиваций для выбора языка имени автора в заголовке библиографической записи:

- на том языке, на котором автор преимущественно публикует свои работы,
- на языке впервые поступившего издания
- на языке страны происхождения автора

- на языке документа

При этом также предлагаются варианты транслитерирования одной языковой графики в другую. При таком многообразии предложений возникает множество вариантов имени одного и того же лица, а пользователь обращаясь при поиске автора к различным каталогам должен каким-то образом их определить и найти. Так, при поиске работ русского автора в каталоге библиотеки Конгресса США, русский пользователь должен сформулировать свой запрос в латинской графике, при этом возможно придется испробовать несколько вариантов написания этого имени в другой графике. Оптимальный путь решения этой проблемы в создании многоязычного файла, когда пользователь зная имя автора, указанное на документе, мог найти его в этом файле и провести дальнейший поиск по предлагаемой модели. Таким образом, необходимо договориться, чтобы имя автора на языке и графике документа, обязательно было приведено в сведениях об ответственности и в авторитетном файле имен. В таком случае выбор языка и графики имени автора в заголовке может быть оставлен на усмотрение библиографирующего учреждения, но более приоритетной мотивацией можно считать – язык национального пользователя и язык страны происхождения автора. Может быть приведена транслитерированная форма имени, так как рекомендуют Англо-американские правила каталогизации для кириллической графики, возможно конверсирование и других языковых график (китайской, японской и т.д.), но обязательным условием должно быть соблюдение соответствующих международных стандартов по транслитерированию различных языковых график. Например, библиографические записи на переводы произведений русского поэта Пушкина А.С., изданные на иностранных языках, в российских каталогах часто приводят в русскоязычной форме имени, но, при этом желательно зафиксировать другие языковые формы в авторитетном файле имен. Кроме того, форма имени автора на языке документа приведена в сведениях об ответственности, где эти условия должны соблюдаться однозначно. Аналогичная позиция сохраняется в отношении языка наименования коллектива в заголовке библиографической записи.

5.2. Uniform title heading.

2.2. Унифицированное заглавие анонимного классического произведения в заголовке библиографической записи.

As a rule a uniform title in Russian rules is an established form of a title of an anonymous classic work with a known national origin and known by users of that nationality or country in their language and script, and also the original language of the work.

Это, как правило, устоявшаяся форма названия произведения, не имеющего автора, однако, имеющего национальное происхождение и потому известного национальному пользователю именно на языке страны происхождения данного анонимного классического произведения.

The most famous Russian anonymous classic work is “Slovo o polku Igoreve.” Russian users know it under the Russian and Cyrillic title though there are versions in other languages. For example, in French it is named “Le Dit de la campagne d’Igor” that is slightly different from the Russian title but still is a version of the original work in Russian. But Russian users typically would not know how to formulate exactly the French uniform title. Moreover, a French title form may be found as transcribed in a title statement where it will be given as the title proper in the language and script of the publication (say in this case in French).

Наиболее известное российское анонимное классическое произведение – «Слово о полку Игореве», российский пользователь знает на русском языке, хотя имеется и другие языковые версии. Например, это произведение, изданное во Франции на французском языке будет называться «Le Dit de la campagne d’Igor». Дословный перевод на русский язык немного другой – «Слово о походе Игоря», но все же это версия с оригинального произведения на русском языке. Да и российский пользователь не сможет точно сформулировать французскую версию унифицированного заглавия

May 18, 2003

заглавия, кроме того французское название можно найти в области основного заглавия, где эти сведения приведены на языке документа.

5.3. Geographic name in a bibliographic record heading.

2.3. Географическое название в заголовке библиографической записи.

In Russian rules language and script for this data corresponds to the language of the imprint data, but this name in the local user's language may be recommended as an additional access point in a bibliographic record or to be included in the authority record in an authority file of geographic names. The complexity of these data is observed through the variety of geographic name forms in different countries and rather frequent changing of names of countries, settlements, etc. That's why the creation of multilingual authority files for geographic names is a very urgent issue now. The Russian State Library has an intention to start creating multilingual authority file of geographic names.

Язык этих данных в заголовке библиографической записи соответствует языку выходных данных документа, но в качестве дополнительной точки в формате или нормативном файле географических названий можно рекомендовать приведение этой же формы географического названия на языке национального пользователя. Учитывая, что географические названия в различных странах мира существенно отличаются друг от друга, а также они имеют тенденцию смены названий (территорий, стран, городов и так далее), доступ к этой информации значительно усложняется. В связи с этим проблема создания и поддержки многоязычного авторитетного файла географических названий стоит более чем актуально. Российская государственная библиотека планирует осуществление проекта по созданию многоязычного файла географических названий.

6. Summary.

3. Заключение.

The international cataloguing rules should provide a more detailed algorithm of choosing bibliographic record language and script, exactly giving a rule which prescribes that main access points should be formulated at least in the language and script of a document's imprint data. We propose that data in languages different from that of a document, for example, transliterated data, may be formulated as additional access points in a MARC record (authority or bibliographic record) according to ISO international standards. Interests of local users may be implemented in notes, bibliographic record headings, subject headings represented in the language of cataloguing and also in a multilingual authority file of names, uniform titles, etc. If all countries follow these international rules, the existing diversity of decisions on choosing bibliographic record language and script in different countries may be overcome, and it will harmonize access to bibliographic information both for local and foreign users. The electronic environment in which an online public access catalogue is being created and functions gives a number of potential alternatives and various possibilities to display a bibliographic record either in a the document's language or the local user's language.

Международные правила каталогизации должны давать более однозначный алгоритм действий в отношении языка и графики библиографической записи, а именно предписывать правило, по которым основные точки доступа в максимальной степени должны быть приведены на языке, объективно отражающем язык и графику выходных данных документа. Формирование данных на других языках, в другой графике отличных от языка документа, например, транслитерированная форма языка реализуется в структуре MARCформата в качестве дополнительных точек доступа по правилам, предписанным международными стандартами ISO. Интересы национального пользователя реализуются в языке примечаний, заголовке библиографической записи, и в полях предметного входа или содержательного блока полей MARCформата, а также на уровне многоязычного авторитетного файла различных данных – имен лиц,

May 18, 2003

наименований коллективов, унифицированных заглавий, географических названий и т.д. Однозначное выполнение международных правил всеми странами позволит преодолеть существующий разрыв в подходах различных национальных правил каталогизации к выбору языка и графики библиографической записи и тем самым гармонизировать доступ к библиографической информации в интересах национального и интернационального пользователя, а электронная среда, в которой формируется и функционирует интерактивный электронный каталог, предоставляет альтернативные или многовариантные возможности для проявления двойственного свойства библиографической записи в коммуникации «язык документа- язык пользователя».