



# 65th IFLA Council and General Conference

## August 20-28, 1999

---

**Code Number:** 099-155(WS)-R  
**Division Number:** IV  
**Professional Group:** Cataloguing: Workshop  
**Joint Meeting with:** -  
**Meeting Number:** 155  
**Simultaneous Interpretation:** -

**Заглавие доклада: ISO 639-1 и ISO 639-2: международные стандарты для кодов языков. ISO 15924: международный стандарт для наименований шрифтов**

**Джон Д. Байрем**  
Библиотека Конгресса США

(John D. Byrum )

---

### **Abstract:**

*Автор описывает два международных стандарта на представление наименований языков. Первый (ISO 639-1), опубликованный в 1988 г., разработан как двухбуквенный код для обозначения 136 языков и применяется в основном в сфере терминологической деятельности. Второй стандарт (ISO 639-2) разработан в конце 1998 г. как трехбуквенный код для обозначения 460 языков. Этот список отвечает потребностям как терминологической, так и библиографической деятельности. Именно поэтому 639-2 рассматривается здесь очень подробно. Появляются его характеристики, излагаются принципы и подходы, определившие работу по составлению списка. Кроме того, автор описывает механизм управления дальнейшей работой над стандартами и их применением. Кратко излагается также содержание текущего проекта по разработке кодов для названий алфавитов, итогом которого будет публикация стандарта ISO 15924. В заключение подчеркивается, что "создание международных стандартов на коды языков и международного стандарта на коды алфавитов является значительным вкладом в достижение Универсального библиографического учета, так как эти списки кодов позволяют передавать и совместно использовать, причем точно, эффективно и на международном уровне, важную информацию о характере публикаций, представленных библиографическими машиночитаемыми записями".*

---

Международная организация по стандартизации (ИСО) занималась разработкой кодовых обозначений языков уже в течение длительного времени. Проект по подготовке стандарта на двухбуквенные коды (далее будут упоминаться также как коды “альфа-2”) был начат несколько десятилетий назад, хотя стандарт ISO 639 (далее ISO 639-1) был опубликован лишь в 1988г. [i]. В том же году началась работа над подготовкой стандарта на трехбуквенные коды (далее будут упоминаться также как коды “альфа-3”), но понадобилось еще 10 лет, чтобы завершить работу над стандартом ISO 639-2: *Коды для представления наименований языков: коды альфа-3* [ii]. Тем временем в середине 90-х гг. началась работа по пересмотру ISO 639-1. Итогом этого начинания должен быть проект международного стандарта.

Технический комитет ИСО 37 “Терминология”, Подкомитет 3 “Оформление словарей”, отвечал за подготовку ISO 639-1. В результате составленный список кодов был ориентирован прежде всего на использование в терминологии, лексикографии и языкознании. ISO 639-1 содержит список кодов для обозначения такого же количества языков. Набор кодов альфа-2 предназначался для практического использования и включал коды наиболее распространенных языков мира, которые чаще всего представлены в мировой литературе. Дополнительные коды создаются в тех случаях, когда становится очевидным, что существует заметный объем литературы на том или ином языке. Индивидуальные коды основаны на оригинальном наименовании языка (если язык пользуется латинской графикой) или конвертированном с помощью латинской графики, за исключением тех случаев, когда соответствующая национальная организация по стандартизации запросила коды, основанные на английской форме наименований языков. Например, японский язык представлен в ISO 639-1 как “ja”.

Как отмечается во введении к стандарту языковых кодов альфа-2, наиболее вероятное использование этих кодов в терминологии и языкознании следующее: 1) для обозначения используемого языка, например, в верхнем поле документа или в библиографических указателях и 2) для обозначения языка, на котором приводится какой-либо термин, например в документах по терминологии, словарях, многоязычных упорядоченных по алфавиту списках слов. В данном стандарте нет прямого указания на использование кодов языков в библиографической деятельности.

ISO 639-2 явился результатом совместной работы, в которой принимали участие представители Технического комитета ИСО 46 “Информация и документация”, Подкомитет 4 “Применение вычислительной техники в информации и документации”. Соответственно, этот стандарт предусматривает не только использование в области лингвистики, но и в контексте гораздо более широкого спектра наименований языков; в ISO 639-2 оговаривается возможность использования кодов языков в библиотеках, информационных службах, в издательском деле для указания языка при обмене информацией, особенно в автоматизированных системах.

Во введении к ISO 639-2 отмечается, что его применение в конкретных случаях было значительно расширено. В дополнение к установленным случаям использования в области терминологии и лингвистики, указанным в ISO 639-1, в списке альфа-3 подчеркивается, что языковые коды пригодны и для передачи библиографической информации. К таким случаям использования относятся указание языков, на которых документы были написаны или введены, например в поле 101 формата UNIMARC для обозначения языка единицы описания, или указание на языки, на которых создаются записи о документах (запись заказа на документ, библиографические записи и т.п.), например в формате UNIMARC, поле 100 “Язык каталогизации”, позиции 22 - 24.

В контексте темы нашего рабочего совещания - “Универсальный библиографический учет в многоязычной среде” - ISO 639-2 с его акцентом на библиографическое применение и расширенным списком языков будет в большей мере, чем ISO 639-1, соответствовать нуждам тех, кто создает и заимствует описания документов всех типов. Именно по этой причине в дальнейшем мое сообщение будет сосредоточено на списке кодов альфа-3.

В ISO 639-2 представлены все языки, отраженные в ISO 639-1, и в дополнение к этому - много других языков и языковых групп, а также коды для особых случаев. В настоящее время языки, представленные в ISO 639-1, являются подмножеством языков, охваченных ISO 639-2; каждый код языка в наборе двухбуквенных кодов имеет соответствующий языковой код в списке альфа-3 и, разумеется, наоборот. В ISO 639-2 насчитывается более 460 кодов. (Языки, разработанные для машинного применения, например языки программирования, не включены ни в тот, ни в другой список.)

Совместная рабочая группа (далее для краткости РГ), которая создала ISO 639-2, приняла решение еще на первой стадии проекта сделать коды в ISO 639-2 совместимыми с ISO 639-1 в той степени, насколько это будет практично. Однако в процессе разработки стандарта возникли значительные трудности с выбором кодов, так как библиографическое сообщество уже располагало широко используемым списком, основанным на списке языковых кодов MARC 21 [iii], который не всегда совпадал с ISO 639-1. Как необходимый компромисс между нуждами терминологии и библиографии (в библиографической деятельности эти коды использовались уже в течение многих лет в сотнях миллионов библиографических записей) РГ согласилась стандартизовать два набора кодов - один для библиографической деятельности (ISO 639-2/B) и один для применения в терминологии (ISO 639-2/T). Оба набора отличаются один от другого в 23 языковых кодах.

Набор кодов В отвечает целям библиографической деятельности и обеспечивает возможность уникального распознавания индивидуальных языков и языковых групп, при этом нет обязательной зависимости от наименования языка, так как коды не построены как аббревиатура наименования языка. Учитывая широкое использование в библиографических записях языковых кодов MARC 21, при разработке набора кодов В был в основном заимствован подход, определивший список кодов в формате MARC. Таким образом, РГ, при разработке списка кодов В, придерживалась следующего критерия для выбора формы кода языка (обычно, но не обязательно, в этом порядке):

- форма, предпочитаемая странами, пользующимися данным языком;
- принятое использование кодов в национальных и международных базах данных;
- национальная или английская форма языка.

Поскольку предполагалось, что ISO 639-2 будет служить и терминологическим нуждам, набор кодов Т основывался на следующем:

- национальная форма языка, или
- предпочтение стран, пользующихся данным языком.

Несмотря на различия в критериях отбора при составлении набора кодов В и набора кодов Т, повторяю, только 23 языковых кода из более чем 460 неидентичны в обоих наборах. Далее, РГ решила, что разработка кодов языков в будущем должна основываться, по мере возможности, на национальной форме наименования языка, за исключением тех случаев, когда от страны или стран, пользующихся этим языком, поступает запрос на иную форму. Сближение наборов кодов В и Т заняло около 10 лет, и результатом явился стандарт ISO 639-2, причем очень часто возникала необходимость напоминать РГ о том, что главная цель создания стандартизованных списков кодов

языков заключается не в том, чтобы стандартизовать наименования языков, представленные кодами, а стандартизовать символы. Важно отметить, что представители библиографического сообщества в составе РГ пошли на многие уступки, чтобы достичь компромисса и успешно завершить проект. В результате предстоит изменить 25 кодов из списка в MARC 21, добавить 33 новых языковых кода и один будет считаться устаревшим. Последствия этих изменений для крупных библиографических баз данных вызывают озабоченность, даже если подлежащие изменению коды представляют относительно незначительные языки (в зависимости от доли литературы на этих языках в фонде какой-либо библиотеки). Однако предполагается, что в будущем список MARC и ISO 639-2 останутся совместимыми.

Принимая во внимание сосуществование двух наборов кодов альфа-3, если для использования выбран какой-нибудь один из наборов, его следует придерживаться во всей его целостности, и о сделанном выборе партнеры по обмену должны быть оповещены до того, как начнется обмен информацией. Исключается возможность заимствования того или иного кода из одного набора, пользуясь кодами из другого. РГ решила также включить специальное положение о том, что коды могут быть изменены лишь в силу неустранимых причин, и если такая замена произошла, любыми прежними кодами нельзя будет пользоваться в течение по крайней мере пяти лет. Еще один принцип, относящийся непосредственно к набору В, сводится к тому, что коды в ISO 639-2 В не будут изменены, чтобы сократить работу по реорганизации списка, даже если изменилось наименование языка, как произошло, например, когда Gallegan стал называться Galician, а Langue d'oc - Occitan.

В ISO 639-2 заложены также некоторые характеристики, которых нет в списке альфа-2. К ним относится положение о "коллективных кодах языков", которые используются, когда объем литературы относительно невелик. Чтобы иметь право на присвоение своего языкового кода, массив литературы должен насчитывать не менее 50 уникальных заглавий и храниться в одном учреждении или в пяти, объединенных в кооператив. В целом массив может включать заглавия на любых носителях, а не только публикации в традиционной печатной форме. Конкретные языки, включенные в каждый коллективный код, не выделены специально в ISO 639-2, как это сделано в списке MARC 21, по решению РГ. Другой специфической характеристикой списка альфа-3 является включение кода (mul) для использования в записях на произведения, которые состоят из нескольких частей на различных языках, и кода (und) для использования в тех случаях, когда необходимо ввести код языка, но наименование языка неизвестно лицу, составляющему запись.

Единый код языка обычно присваивается языку, пользующемуся несколькими видами письма, например Sindhi, на котором пишут арабским письмом, а также письмом Gurmukhi и Devangari, или Somali, на котором пишут арабским и латинским письмом. Как исключение, отдельные коды предусмотрены для хорватского и сербского языков, хотя многие эксперты считают, что здесь мы имеем случай с одним и тем же языком, на котором одни пишут латинским алфавитом, а другие - кириллическим.

Обычно все диалекты какого-нибудь языка представлены кодом для этого языка, но в ряде случаев (в силу исторических обстоятельств, а не как принцип) имеются коды и для диалектов, например в случае с Awadhi, диалектом хинди. ISO 639-2 признает, что в определенных случаях учреждению могут понадобиться коды для диалектов; поэтому были зарезервированы коды от qaa до qtx для местного использования. Другим примером, когда требуются местные коды, могут быть коды для древних языков, которые в ISO 639-2 не имеют своих кодов. Но в стандарте оговаривается, что записи с кодами, предназначенными для местного использования, не подлежат международному обмену, так как местные коды будут в каждом учреждении разными.

Безусловно, ведение и соблюдение международного стандарта очень важны для обеспечения его соответствия изменяющимся условиям и требованиям. В случае с ISO 639 были назначены два регистрационных органа. Для списка альфа-2 это Infoterm со штаб-квартирой в Вене, Австрия, а для списка альфа-3 - Библиотека Конгресса США в Вашингтоне. Оба органа отвечают за получение и рассмотрение предложений о введении новых или изменении существующих кодов языков. Как уже упоминалось выше, запросы на введение дополнительных кодов, не представленных среди языков в ISO 639-2, должны быть обоснованы наличием 50 заглавий. Когда отклоняется запрос на новый код, он может быть зарезервирован для использования заявителем и для других потенциальных пользователей. В каждом случае регистрационные органы рекомендуют дальнейшие действия Совместному консультативному комитету (Joint Advisory Committee - JAC), который наблюдает за поддержанием стандарта в целом. В JAC соблюдается равное представительство Технических комитетов ИСО 37 и 46, а председатель переизбирается каждые два года, представляя соответственно Infoterm и Библиотеку Конгресса. Принимая решения, члены JAC должны единогласно одобрить предложения при первом голосовании. Если это невозможно, необходимо второе голосование, причем должно быть не менее пяти голосов "за". Равное представительство и строгие требования к единогласию способствуют обеспечению надежности будущего ISO 639-1.

Предположительно JAC соберется в октябре 1999 г. для обсуждения технических вопросов, возникших на стадии сбора комментариев, когда ISO 639-2 получил статус проекта международного стандарта. Более общий вопрос, подлежащий обсуждению, - это будущие взаимоотношения между языковыми кодами альфа-3 и теми, которые вошли в список альфа-2, включающий ISO 639-1, который сейчас находится в процессе пересмотра.

В настоящее время ИСО осуществляет еще один проект, имеющий отношение к данному рабочему совещанию: Технический комитет 46, Подкомитет 2 "Преобразование письменных языков" работает над составлением списка кодов для представления наименований алфавитов (шрифтов) [iv]. Эти коды, так же как и описанные выше коды языков, предназначаются для использования в терминологии, лексикографии и языкознании и в других случаях, когда требуется выразить наименование алфавита (шрифта) в кодированной форме, включая машинную обработку в библиографических целях. В разрабатываемом стандарте предусмотрены три кода, а не один, для каждого названия алфавита (шрифта): 1) двухбуквенный код и 2) трехбуквенный код, образуемые обычно из оригинального названия шрифта на языке, в котором он используется, транслитерированного или транскрибированного буквами латинского алфавита, а также 3) числовой вариант, разработанный, чтобы "в какой-то мере обеспечить мнемоничность используемых кодов". Для числовых кодов был разработан диапазон цифр, позволяющий охватить характер шрифта (письма), например: 000-099 - для иероглифов и клинописи, 100-199 - для письма справа налево, 200-299 - для письма слева направо и т.д. Диапазон 700-899 не присваивается, тогда как 900-999 зарезервирован "для частных случаев использования, как альтернативные формы для многочисленных шрифтов и специальных кодов". Коды для алфавитных письменностей описаны согласно ISO 639-1 и ISO 639-2 без указаний на предпочтительное использование в терминологической или библиографической деятельности, содержащихся в ISO 639-2. В настоящий проект включено 95 - 100 наименований шрифтов и их альтернативных наименований. Когда он будет утвержден как международный стандарт, за его ведение и поддержание будет отвечать соответствующий регистрационный орган. В заключение необходимо отметить, что создание международных стандартов на коды языков и международного стандарта на коды алфавитов явится значительным вкладом в достижение Универсального библиографического учета, так как эти списки кодов позволят передавать и

совместно использовать, причем точно, эффективно и на международном уровне, важную информацию о характере публикаций, представленных библиографическими машиночитаемыми записями.

### **Примечания**

- i. ISO 639: Code for the representation of names of languages / 1st ed. - Geneva: Intern. Standardization Organization, 1988. - 17 p.
- ii. ISO 639-2: Codes for the representation of names of languages: Alpha-3/ 1st ed. - Geneva: Intern. Standardization Organization: 1998. - 66 p.
- iii. MARC 21: название формата, созданного в результате гармонизации форматов USMARC и CAN/MARC . Опубликовано в 1999 г.
- iv. ISO 15924: Code for the representation of names of scripts. - 18 p. Проект, датируется 9 июля 1998 г.