

ALA-LC
ROMANIZATION
TABLES

Transliteration Schemes for Non-Roman Scripts

Library of Congress

Cataloging Distribution Service

ALA-LC ROMANIZATION TABLES

Transliteration Schemes for Non-Roman Scripts

Approved by the
LIBRARY OF CONGRESS
and the
AMERICAN LIBRARY ASSOCIATION

Tables compiled and edited by
RANDALL K. BARRY
Network Development and MARC Standards Office

Library of Congress ■ Washington ■ 1991
Cataloging Distribution Service

411
Ab1
cop.2

General Slavic Ref. PR
(Slavic)

CONTENTS

INTRODUCTION v

ROMANIZATION TABLES (arranged alphabetically) 1

CHARACTER SETS USED IN ROMANIZATION 210

INDEX TO LANGUAGES 213

25

INTRODUCTION

ALA-LC Romanization Tables is a new collection of 50 romanization and transliteration schemes, the majority of which were previously published in the Library of Congress *Cataloging Service Bulletin*. These tables were developed for use when the consistent transliteration of a non-Roman (vernacular) script (e.g., Arabic script) into the Roman alphabet is needed. This publication covers more than 140 languages written in various non-Roman scripts.

Transliteration of non-Roman scripts is done by librarians, archivists, information specialists, writers, etc. when translation into English or some other Roman script language is not desirable or possible. This is often the case with proper names, titles, and terms for which no appropriate Roman script equivalent exists. For some scripts the process of romanization or transliteration involves imparting characteristics of the Roman script such as capitalization and word division. The Thai script is typical of one that has neither of these characteristics. (There are no separate upper and lower case forms for letters in the Thai script and words are not separated by spaces.) *ALA-LC Romanization Tables* provides special guidelines for capitalization and word division for certain scripts.

Development of ALA-LC Romanization Tables

Romanization schemes have been developed over the years by foreign language specialists at Library of Congress, often with the help of scholars and experts at other institutions. This work is done to support the cataloging of foreign language materials. Bibliographic information appearing in vernacular scripts is routinely romanized so that it can be incorporated into the Library's large catalog and shared with other institutions. Cataloging information in the vernacular is often included with the romanization.

As tables for new scripts are developed they are reviewed and approved by the American Library Association. This provides input from a wider group of potential users. The development of romanization schemes is an ongoing process. Periodically new romanization schemes are developed for scripts for which no scheme exists. Existing schemes are occasionally revised to reflect better understanding of a script or to provide an improved rendering of it in the Roman script.

Changes in this Edition

The tables included in this edition supersede all of the ALA-LC romanization tables issued previously. All of the tables have been reformatted to improve legibility and correct any typographical errors found in the originals. The handwritten portions of tables for several scripts have been replaced by modern print fonts. The presentation of information relating to romanization, capitalization, and word division was made more consistent. Instructions regarding the encoding of special characters (e.g., apostrophe) and character modifiers (i.e., diacritics) in machine-readable form have also been added. This information is especially important for librarians who create records following the USMARC formats.

Two new tables, for Komi (Molodtsov) and Tigrinya, are appearing for the first time to cover non-Roman scripts for which no ALA-LC table has ever been published. The word division guidelines for Korean have been completely revised. A special section on character sets has been added as well as a useful index to languages covered by this publication.

Special Features of the Tables

For each non-Roman script, a font was selected that presents each character (i.e., alphabetic, syllabic, ideographic, etc.) as unambiguously as possible. For some scripts certain characters are stylized. This was done to make similar characters more easily distinguished from one another.

The organization of characters belonging to each script follows the traditional arrangement of characters for the script. Often, this is not in the order of the Roman alphabet.

For scripts where certain characters and character modifiers can appear in combination with others (e.g., vowel marks written above or below other characters), an unspecified character is represented by the convention "□" (e.g., □ represents the acute accent which can appear over a variety of alphabetic characters). This convention is only used for characters which are normally written in combination with other characters. These characters are referred to as *character modifiers* and are generally considered *non-spacing*.

Most scripts included special characters that are used in addition to the regular letters of an alphabet, syllabary, or system of ideographs. These special characters occupy a space by themselves (e.g., ' *apostrophe*). When presented in one of these romanization tables, they are not shown in conjunction with any □.

The romanization shown for each character often includes one or more special characters or character modifiers. Special attention should be paid to this part of each romanization scheme. For characters that might be misidentified, a special legend (e.g., (*soft sign*)) is included in parentheses following the character. When a character has no roman value this is indicated by the presence of a hyphen (-), the word "*omit*", or "*disregard*" in the romanization table.

Examples have been included of vernacular text for some scripts. These examples are intended to be descriptive of the rule or principle presented.

NOTE: In the application of the *ALA-LC Romanization Tables*, familiarity with the language involved is helpful, if not essential, for many scripts.

List of Romanization Tables

<i>Language</i>	<i>page</i>
Amharic	2
Arabic	4
Armenian	14
Assamese	16
✓ Belorussian	18
Bengali	20
✓ Bulgarian	22
Burmese	24
Chinese	27
✓ Church Slavic	40
Divehi	43
✓ Georgian	46
Greek	48
Gujarati	50
Hebrew and Yiddish	52
Hindi	54
Japanese	57
Kannada	70
Kashmiri	72
Khmer	76
Korean	79
✓ Kurdish	94
Lao	96
Lepcha	100
Limbu	102
Malayalam	104
Marathi	106
Mongolian	108
Moplah	110
✓ Non-Slavic Languages (in Cyrillic Script)	114
Oriya	132
✓ Ottoman Turkish	134
Pali	140
Persian	145
Punjabi	152
Pushto	154
✓ Russian	160
Sanskrit	162
Santali	164
✓ Serbian and Macedonian	166
Sindhi	168
Sinhalese	170
Tamil	172
Telugu	174
Thai	176
Tibetan	192
Tigrinya	194
Uighur	196
✓ Ukrainian	200
Urdu	202