

## International Cataloging: Use Non-Latin Scripts

Revised: September 2011



6565 Kilgour Place, Dublin, OH 43017-3395  
[www.oclc.org](http://www.oclc.org)

## Revision History

Date	Section title	Description of changes
September 2011	1 Connexion client international cataloging	Updated the section about adding input languages and keyboards provided by Windows for entering non-Latin-script data, including two example procedures, one for Windows 7 and one for Windows XP.
January 2008	1 Connexion client international cataloging	Added a note about selecting an option to receive a warning before exporting records with parallel unlinked non-Latin script fields.
	2 Use non-Latin scripts for cataloging	<ul style="list-style-type: none"> <li>• Added same note as above.</li> <li>• Added a list of script codes to use with the WorldCat “character sets present” index (label <b>vp:</b>) to retrieve all records that contain specified script data.</li> </ul>
	3 Add non-Latin script variant name headings in authority records	New topic for upcoming support for using MARC-8 character sets to add non-Latin script data in variant name heading and some note fields of authority records. Made available by the Library of Congress and its partners after April 1, 2008.
	All topics on specific scripts (sections 4 -6 and 8 - 13)	Added directions for retrieving all records containing the particular script covered by the topic (see “Notes on searching”).

© 2011 OCLC

The following OCLC product, service and business names are trademarks or service marks of OCLC, Inc.: Connexion, OCLC, WorldCat, and "The world's libraries. Connected."

In addition, the WorldCat symbol and OCLC symbol are service marks of OCLC. Third-party product and service names are trademarks or service marks of their respective owners.

OCLC grants permission to photocopy this publication as needed.

## Table of Contents

### **1 Connexion client international cataloging**

- Scripts and languages supported, 6
- Scripts supported for variant name headings in authority records, 6
- Valid character sets for supported scripts, 6
- Guidelines for contributing non-Latin script bibliographic records to WorldCat, 8
- How the client manages non-Latin script data in bibliographic records, 9
- Windows provides input languages for using non-Latin scripts, 10
- Change the language of the client interface, 12
- Summary of general international features, 12
- Client tools specifically for using non-Latin scripts, 13
- Non-Latin script bibliographic records in Connexion browser, 14

### **2 Use non-Latin scripts for cataloging bibliographic records**

- Basic client cataloging functions and non-Latin scripts, 15
- Search WorldCat, 15
- Display non-Latin scripts in records and lists, 18
- Create records, 18
- Edit records, 20
- Use constant data, 24
- Save records and search save files, 24
- Report errors in non-Latin script records, 25
- Export records, 25
- Import records, 27
- Using non-Latin script data in macros, 27

### **3 Add non-Latin script variant name headings in authority records**

- About using non-Latin scripts for variant name headings in LC authority file records, 28
- Details, 28
- Character sets supported, 29
- Existing non-Latin script support in Connexion client for bibliographic records:  
What applies to authority records?, 29

### **4 Catalog using Arabic scripts**

- About using Arabic scripts, 32
- Tools for using non-Latin scripts, 32
- Arabic script entry and character sets, 33
- Transliterate romanized data in Arabic or Persian records into Arabic script, 34
- Results of transliteration and auto-transliteration, 35
- Basis of transliteration, 36
- Align Arabic or Hebrew script data for display and print, 36
- Use Unicode formatting characters to control bidirectional data, 36
- Use Arabic definite article in Arabic script searches, 38
- Indexing for Arabic script searches, 38

### **5 Catalog using Bengali script**

- About using Bengali script, 43
- Tools for using non-Latin scripts, 43
- Unicode export and import required for Bengali records, 43
- About Unicode, 43
- Bengali script entry and character set, 44
- Indexing for Bengali script searches, 44

- 6 Catalog using Chinese, Japanese, and Korean (CJK) scripts**
  - About using CJK scripts, 46
  - Tools for using non-Latin scripts, 46
  - CJK entry and character set, 47
  - CJK E-Dictionary, 48
  - Convert invalid CJK characters to equivalent MARC-8 characters, 48
  - Use the Chinese, Japanese, or Korean client interface, 48
  - Indexing for CJK script searches, 49
- 7 Use CJK E-Dictionary**
  - What is the CJK E-Dictionary?, 51
  - Why search or browse the CJK dictionary?, 51
  - How to search or browse the CJK E-Dictionary, 52
  - EACC layers: details, 54
  - Use the search results list, 55
  - Open and use a dictionary entry, 57
  - What is EACC?, 57
  - What is Unicode?, 58
  - What is Tsang-chieh?, 59
- 8 Catalog using Cyrillic scripts**
  - About using Cyrillic scripts, 64
  - Tools for using non-Latin scripts, 64
  - Cyrillic script entry and character set, 65
  - Indexing for Cyrillic script searches, 65
- 9 Catalog using Devanagari script**
  - About using the Devanagari script, 67
  - Tools for using non-Latin scripts, 67
  - UTF-8 Unicode export and import required for Devanagari records, 67
  - About Unicode, 68
  - Devanagari script entry and character set, 68
  - Indexing for Devanagari script searches, 68
- 10 Catalog using Greek script**
  - About using Greek script, 70
  - Tools for using non-Latin scripts, 70
  - Greek script entry and character set, 70
  - Indexing for Greek script searches, 71
- 11 Catalog using Hebrew script**
  - About using Hebrew script, 73
  - Tools for using non-Latin scripts, 73
  - Hebrew script entry and character set, 74
  - Align Hebrew or Arabic script data for display and print, 75
  - Use Unicode formatting characters to control bidirectional data, 75
  - Indexing for Hebrew script searches, 77
- 12 Catalog using Tamil script**
  - About using Tamil script, 79
  - Tools for using non-Latin scripts, 79
  - Unicode export and import required for Tamil records, 79
  - About Unicode, 79
  - Tamil script entry and character set, 80
  - Indexing for Tamil searches, 81

### **13 Catalog using Thai script**

- About using Thai script, 83
- Tools for using non-Latin scripts, 83
- Unicode export and import required for Thai records, 83
- About Unicode, 83
- Thai script entry and character set, 84
- Indexing for Thai script searches, 84

### **14 Install RLIN21 Arabic, Cyrillic, Hebrew or Latin keyboards**

- About keyboards for entering script data in the Connexion client, 87
- Install an RLIN21 keyboard, 87
- Make an installed RLIN21 keyboard available for use, 88
- Select an RLIN21 keyboard to input script data, 89
- Uninstall an RLIN21 keyboard, 90
- Guides - RLIN21 keyboard layouts and RLIN21 Latin character keystroke shortcuts, 91

## 1 Connexion client international cataloging

**Scripts and languages supported** Connexion client supports the following non-Latin scripts for cataloging items in languages that use the scripts:

Script	Examples of supported languages
Arabic	Arabic, Persian Urdu, Azerbaijani
Bengali	Bangla, Assamese
Chinese	Chinese
Cyrillic	Russian, Bulgarian, Serbian, Ukrainian
Devanagari	Hindi, Marathi, Sanskrit, Nepali, Sherpa
Greek	Greek
Hebrew	Hebrew
Japanese	Japanese
Korean	Korean
Tamil	Tamil
Thai	Thai

**Note:** You can include more than one non-Latin script anywhere in a record, including within the same field.

**Scripts supported for variant name headings in authority records** Only MARC-8 character sets will be supported for adding non-Latin script variant name headings in authority records. The following scripts are supported:

- Arabic (including the Persian language)
- Chinese
- Cyrillic
- Greek
- Hebrew (including Yiddish)
- Japanese
- Korean

See "Add non-Latin script variant name headings in authority records" in this Guide for more information.

**Valid character sets for supported scripts** **Arabic, CJK, Cyrillic, Greek, and Hebrew**

Character sets for these scripts given in *MARC 21 Specifications for Record Structure, Character Sets, and Exchange Media* on the Library of Congress Web site at: <http://www.loc.gov/marc/specifications/spechome.html> define the scope of valid characters in Connexion client. The MARC-8 character set is the subset of Unicode characters approved for use in MARC 21 cataloging.

The following list defines the scope of valid characters in the Connexion client for Arabic (including Persian), CJK, Cyrillic, Greek, and Hebrew scripts:

- 33(hex) [ASCII graphic: **3**] Basic Arabic
- 34(hex) [ASCII graphic: **4**] Extended Arabic
- 31(hex) [ASCII graphic: **1**] Chinese, Japanese, Korean (EACC)
- 4E(hex) [ASCII graphic: **N**] = Basic Cyrillic
- 51(hex) [ASCII graphic: **Q**] = Extended Cyrillic
- 53(hex) [ASCII graphic: **S**] = Basic Greek
- 32(hex) [ASCII graphic: **2**] = Basic Hebrew

**Note:** The client inserts the notation (**3**, (**4**, **\$1**, (**N**, (**Q**, (**S**, or (**2**, respectively, into field 066 to indicate which script(s) are used in a record. If multiple scripts are used, the notations are inserted individually, each in a separate subfield c.

### **Bengali, Devanagari, Tamil, and Thai**

There are no MARC-8 character sets for Bengali, Devanagari, Tamil, or Thai. OCLC implemented the following script identification codes for these scripts based on ISO 15924 Code Lists (<http://www.unicode.org/iso15924/codelists.html>).

The following list shows the ranges of UTF-8 Unicode characters that define valid characters for these scripts in the Connexion client:

- **Beng** = Bengali (character range U+0980 to U+09FF)
- **Deva** = Devanagari (character range U+0900 to U+097F)
- **TamI** = Tamil (character range U+0B80 to U+0BFF)
- **Thai** = Thai (character range U+0E00 to U+0E7F)

**Note:** The client inserts **Beng**, **Deva**, **TamI**, or **Thai**, respectively, in field 066 of a record to indicate that the script is used. If multiple scripts are used, the notations are inserted individually, each in a separate subfield c.

### **Limitations on using Bengali, Devanagari, Tamil, and Thai scripts**

- To export or import records containing Bengali, Devanagari, Tamil, and Thai scripts, you must select the UTF-8 Unicode character set option in:

To export:


— **Tools > Options > Export**; click **Record Characteristics** and select **UTF-8 Unicode** in the **Character Set** list under **Bibliographic Records**.

— **To import: File > Import Records**; click **Record Characteristics** and select **UTF-8 Unicode** in the **Character Set** list under **Bibliographic Records**.

- Because Bengali, Devanagari, Tamil, and Thai scripts are not part of MARC-8 characters, you cannot export or import these scripts using the MARC-8 character set option.

- Because MARC-8 characters are part of UTF-8 Unicode, you can safely export or import Arabic, CJK, Cyrillic, Greek, and Hebrew records using either the MARC-8 or the UTF-8 Unicode character set option.
- Bengali, Devanagari, Tamil, and Thai scripts are not supported for the following:
  - Variant name headings in authority records
  - MARC Subscription
  - Bibliographic Record Snapshot

### Invalid characters in Connexion client

Any characters that are not included in the above lists of defined characters or that cannot be inserted via **Edit > Enter Diacritics** (or  or <Ctrl><E>) are **invalid** in the client. To include non-Latin characters that you need but that are invalid in Connexion client, you can:

- Enter the character in the record, export the record to your local system using Unicode export format, and then remove the character before processing the record in WorldCat.

Or

- Enter the name of the character within square brackets, using the Unicode standard if available, (for example, enter **[schwa]**), or for CJK characters, enter the reading of the character (for example, enter **[yin]**).

For reference, see, for example, the Unicode charts Web page at <http://www.unicode.org/charts/>, which has a character name index.

**Note:** Z39.50 access to WorldCat records also supports MARC-8 and Unicode UTF-8 character sets. See information on non-Latin script support in Z39.50 in the Z39.50 cataloging documentation at: <http://www.oclc.org/support/documentation/z3950/searchtips#5>


### Multiscripts in a single record are valid

Use as many supported non-Latin scripts as you need anywhere in a record, including within the same field.

#### Guidelines for contributing non-Latin script bibliographic records to WorldCat

Records added to WorldCat must meet MARC standards, no matter what type of scripts you use to enter the data. You must catalog according to AACR2 practices. For more information, see *OCLC Bibliographic Formats and Standards* on the OCLC Web site at: < <http://www.oclc.org/bibformats/en/about/> >.

For quick and easy reference, open a detailed *Bibliographic Formats* description of any field directly from Connexion client:

Action
Place the cursor in a field and click <b>Help &gt; MARC Field Help</b> (or click  or press <Shift><F1>). Or Right-click in the field, and on the pop-up menu, click <b>MARC Field Help</b> .

**Romanized data.** If you provide romanized (Latin-script-equivalent) data, romanization should follow guidelines in the *ALA - LC Romanization Tables* on the Library of Congress Web site at: < <http://www.loc.gov/catdir/cpsd/roman.html> >.

### How the client manages non-Latin script data in bibliographic records

For cataloging items in languages that use non-Latin scripts, create records that contain:

- Non-Latin script data only (or multi-non-Latin scripts if needed, one script per field)
- Latin-script equivalent data only

Or

- Both non-Latin and Latin scripts

If you include both, the client provides (or you create) paired fields that have the same tag number. The top field of the pair is for non-Latin-script data followed by a field for the corresponding romanized data.

For machine-processing purposes, non-Latin script fields are stored internally in MARC format 880 fields.

How paired fields work:

- **Automatic linking.** The client automatically links two fields when a non-Latin script field is followed by a Latin-script field (romanized, or Latin-script-equivalent, data) that displays the same tag number. The links are added when you reformat, save, or take a final OCLC action on the record.
- **Link and Unlink commands.** Whether paired fields are present or absent in a record, you can link or unlink fields using **Edit > Linking Fields > Link Fields** or **Unlink Fields** (or <Alt><E><K><L> or <Alt><E><K><U>, respectively). The client displays linked fields with a connecting bracket in the left margin.

**Caution:** If Latin script and non-Latin script parallel fields are not linked, display of the non-Latin script in records downloaded to your local system may be affected. You can set an option to get a warning before the client exports records with parallel unlinked non-Latin script fields in **Tools > Options > Export**.

- **Field 245.** If no romanized data appears in field 245, the client automatically adds:
  - A 245 field containing the following three filler characters: < > . (less than bracket, greater than bracket, and a period)
  - A 500 field with the text *Non-Latin script record*
- **ISBN.** If you enter an ISBN into a non-Latin script 020 field without adding it to the paired Latin script 020 field, the client automatically copies the data to the Latin-script 020 field.
- **Automatic 066.** When you validate, reformat, save, or take a final OCLC action (interact with the OCLC system) a non-Latin script record, the client automatically adds the 066 field with the following data in ꜥc to indicate which character set(s) the record contains:
  - (3** for basic Arabic
  - (4** for extended Arabic
  - Beng** for Bengali
  - \$1** for CJK
  - (N** for basic Cyrillic
  - (Q** for extended Cyrillic
  - Deva** for Devanagari
  - (S** for Greek
  - (2** for Hebrew
  - TamI** for Tamil
  - Thai** for Thai

### Windows provides input languages for using non-Latin scripts

If the default language of your workstation is not the language you need for entering supported non-Latin scripts for cataloging or for variant name headings in authority records, or if you do not already have an input method for the language set up on your workstation, install input languages and methods in Windows.

**Note on alternative RLIN21 keyboards:** For information on alternative keyboards available for download for some scripts, see "Install RLIN21 Arabic, Cyrillic, Hebrew or Latin keyboard" in this Guide.

The exact procedure for adding input languages and keyboards will vary, depending on your version of Windows. The following are example procedures for two versions of Windows.

#### Example—In Windows 7:

1. Go to **Start > Settings >; Control Panel > Clock, Language and Region**.
2. Click **Regional and Language**.
3. Click the **Keyboards and Languages** tab, and then click **Change keyboards**.
4. Under Installed services, click **Add**. Double-click the language you want to add and double-click **Keyboard**.
5. Repeat step 4 to add more languages if needed.

### Example—In Windows XP:

1. Go to **Start > Settings > Control Panel > Regional and Language Options**.
2. Click **Languages**.
3. **Optional**. In the **Languages** tab, click a check box to select "Install files for complex script and right-to-left languages (including Thai)."
4. **Required** for creating Chinese, Japanese, and Korean text: Click a check box to select "Install files for East Asian languages."
5. Click **Details**. In the Settings tab, click **Add**.
6. Select a language to install and click to select the **Keyboard layout/IME** check box to also install the keyboard.

When English is your default language and you install one or more other input languages, Windows automatically displays a language bar on the desktop for toggling through your input languages. You can also use the default Windows keystroke shortcut <Left Alt><Shift> to toggle through installed languages.

For information and help on setting up input languages and keyboards specifically for your version of Windows:

- On the Windows **Start** menu, click **Help**, and in the Help Search box, enter "input languages."

Or

- Go to the Microsoft Web site ([www.microsoft.com](http://www.microsoft.com)), and search the site for "input languages vista" or "input languages windows 7," for example.

Please consult with your system administrator if you need help with installing input languages and keyboards.

**Change the language of the client interface**

The client prompts you to select an interface language when you open the client for the first time after installing or when you create a new user profile.

Or

You can change the language of the interface anytime by changing an option:

	Action
1	On the Tools menu, click <b>Options</b> (or press <Alt><T><O>), and then click the <b>International tab</b> .
2	In the <b>Interface Language</b> list, select one of the following available languages: <ul style="list-style-type: none"> <li>• Chinese (Simplified)</li> <li>• Chinese (Traditional)</li> <li>• English (<b>default</b>)</li> <li>• German</li> <li>• Japanese</li> <li>• Korean</li> <li>• Spanish</li> </ul> <p><b>Note:</b> To display the Chinese, Japanese, or Korean interface, you must have an input method for the language installed on your workstation or a Chinese, Japanese, or Korean language version of Windows.</p>
3	Click <b>OK</b> (closes the Options window) or <b>Apply</b> (keeps the window open) to change to the language you selected. Or Click <b>Cancel</b> to close the Options window without changing the interface language. <b>Result when you change the language:</b> <ul style="list-style-type: none"> <li>• Text in the client interface displays immediately in the selected language.</li> <li>• The online client Help is provided in English only.</li> </ul>

**Summary of general international features**

- **Access to records with non-Latin scripts.** Catalogers who use the Connexion client can view, create, edit, and take actions on records with supported non-Latin script data. Non-Latin scripts are not visible in WorldCat records in the Connexion browser interface unless they are saved to the online save file using the Connexion client. Save file records containing non-Latin scripts are read-only in the browser.
- **Details for using non-Latin scripts:**
  - See details in the next section, "Use non-Latin scripts for cataloging."
  - See separate topics in this Guide about using each specific non-Latin script.
- **Interface language of the client.** See the procedure above to change the default language (English) of the client interface to Chinese (Simplified or Traditional), German, Japanese, Korean, or Spanish.

Select the interface language when you open the client for the first time after installing, when you create a new user profile, or anytime in **Tools > Options > International**.

**Note:** If you select a language other than English, only the client interface displays in the selected language. The online client Help is provided in English only.

- **Chinese name authority file.** Search the Chinese name authority file (**Authorities > Search > Chinese Name Authority File**).
  - The Joint University Librarians Advisory Committee (JULAC) of Hong Kong creates and maintains the Chinese name authority file.
  - Anyone using Connexion client with an existing authorization/password can access Chinese name authority records in this file.
  - Use keyword or numeric searches in either the command line (enter full search syntax) or the guided keyword/numeric search area (enter or select search components).
  - Access to Chinese name authority records is read-only. You may copy or print only.  
  
See topics on searching the Chinese name authority file in *Authorities, Search Authority Files* for details.
- **Authority records.** You can add MARC-8 non-Latin script variant name headings to authority records and export for use in your local system. If you are a NACO participant, you can add non-Latin script name headings and replace the master records in the LC authority file.

### Client tools specifically for using non-Latin scripts

The client provides the following tools to help you catalog using non-Latin scripts:

- Export and import character set option - choose MARC-8 (**default**) or UTF-8 Unicode.  
**Notes:**
  - MARC-8 character sets supported for Arabic, CJK, Cyrillic, Greek, and Hebrew scripts are subsets of UTF-8 characters. You can export or import records containing these scripts using either the MARC-8 or UTF-8 character set option.
  - The client also supports UTF-8 Unicode character sets (not included in MARC-8 characters) for Bengali, Devanagari, Tamil, and Thai scripts. **Caution:** You must select the UTF-8 Unicode options to export or import records containing Bengali, Devanagari, Tamil, or Thai scripts.
- Export options for data fields - determine types, position, and sort order of Latin and non-Latin script data for exported records.
- MARC-8 character verification - verify MARC-8 characters separately from record validation (this function is inappropriate for checking Bengali, Devanagari, Tamil, and Thai characters, since they are not included in MARC-8 characters sets; Bengali, Devanagari, Tamil, and Thai characters are verified during record validation).
- Field linking/unlinking - visually link or unlink non-Latin script data fields with equivalent romanized data fields.

- Arabic and Persian transliteration - two ways to automatically transliterate existing romanized data (Latin script representation of non-Latin script) into Arabic script data for Arabic and Persian records.
- Data alignment for displaying and printing Arabic and Hebrew script data.
- Unicode formatting control characters to support correct display of bidirectional data in Arabic and Hebrew script records (use right-click menu).
- CJK E-Dictionary - helps with character selection by providing comprehensive information about CJK characters supported in the client.
- Conversion of invalid CJK to MARC-8 - automatically convert invalid CJK characters to equivalent MARC-8-compliant characters.

### Non-Latin script bibliographic records in Connexion browser

- **WorldCat records.** Non-Latin script data in a WorldCat record does not display when you open the record using the Connexion browser. The message *Non-Latin script suppressed* displays in the upper right of the record. You cannot lock or replace these records.
- **Online save file records.** If you save a record containing non-Latin scripts to the online save file using the Connexion client and then open the record in the Connexion browser:
  - The record opens in display mode only with a warning at the top: *This record contains non-Latin script data and cannot be edited using this interface.* You cannot edit or take final actions on the record.
  - All non-Latin script data is displayed in 880 fields at the end of the record.
  - You can view the record, print it, or copy and paste data from it.
  - You cannot flag records containing non-Latin scripts.

## 2 Use non-Latin scripts for cataloging bibliographic records

**Basic client cataloging functions and non-Latin scripts** This topic describes how the Connexion client supports Arabic, Bengali, CJK, Cyrillic, Devanagari, Greek, Hebrew, Tamil, and Thai scripts by specific cataloging function.

This topic contains the following kinds of information:

- Procedures for using client tools and options specifically developed for non-Latin script records
- Any special parameters for using non-Latin script data with existing client functionality

### Search WorldCat

#### Entering searches

- Search for records containing non-Latin script data using either script search terms or romanized (Latin-script equivalent) search terms.
- Both interactive and batch searching support non-Latin script search terms (**Cataloging > Search > WorldCat** and **Batch > Enter Bibliographic Search Keys**).
- Alternatively, copy and paste non-Latin script data into client searches from sources external to the client.
  - Non-Latin script search terms must be based on Unicode. However, only Unicode characters that can be converted to MARC-8-equivalent characters are valid in WorldCat.
  - If Unicode characters that are not convertible are in the search term, you may find no matching records.
- About using search indexes for non-Latin script search terms:
  - Use the same indexes (labels and punctuation) for non-Latin script searches that you use for Latin script searches. Enter the index labels and punctuation using Latin script.
  - Do not use derived searching for non-Latin scripts.
  - Add the same qualifiers to both Latin and non-Latin script searches. Enter them using Latin script.
  - Truncating searches is supported for Bengali, CJK, Devanagari, Tamil, and Thai script search terms only, not for other non-Latin script searches (for other scripts, use browsing for automatic truncation if needed).

To truncate, end the search term with an asterisk to find the string preceded by the asterisk followed by any other data. You do not have to enter the entire data string as it appears in a field or subfield.
  - Browsing to scan indexes for a match is available for any supported script. Browsing scans for the exact data string followed by any other data, providing automatic truncation. Enter only as many characters as are unique enough to retrieve matching record(s).
  - For **Arabic, Bengali, Cyrillic, Devanagari, Greek, Hebrew, and Tamil** script searches, use word or phrase search indexes and word or phrase browse indexes.

—For **CJK** script searches, the system indexes both single characters and immediately adjacent characters in a field. Use the following search strategies:

- Word search - Enter an index label and a colon (for example **ti:**) followed by a character string with no spaces to find a single word, or followed by more than one character string separated by a space to find multiple words, anywhere in an indexed field.
- Phrase search - Enter an index label and an equal sign (for example, **ti=**) followed by a character string to find exact occurrences, starting with the first character in an indexed field and including each succeeding character. Truncate the character string to find the string followed by any other data without having to enter the entire data string as it appears in a field or subfield.

**Note:** Enter a minimum of three CJK characters if you truncate a search.

- Phrase browse - Enter the Scan command, an index label, and an equal sign (for example, **sca ti=**) followed by a character string. Phrase browsing scans an index for occurrences of the browse string at the beginning of indexed fields, followed by any other data (automatic truncation).

**Note:** Since all MARC-8 CJK characters are indexed singly, if you browsed for a word, the system would scan for the first character only, and results would not be significant.

—For **Thai** script searches, the system treats the entire data string you enter as both a word and a phrase, since Thai text has no spaces between words. Search for Thai terms using word or phrase search indexes and word or phrase browse indexes.

For phrase searching, you must use truncation.

To truncate, enter an asterisk (\*) at the end of the search string. Enter a minimum of three Thai characters before truncating.

**Note on word and phrase searching and browsing in general:** Word searches find the data string you enter anywhere in the indexed field. Phrase searches find the data string starting with the first character in a field or subfield and including every character in exact order. Browsing scans an index for the closest match to the character string followed by any other data.

- If you want to retrieve all records or see sample records containing a particular script, use the "character sets present" WorldCat search index (label **vp:**) with the assigned code for a script:

Script	Code for script	Enter search as ...
Arabic	<b>ara</b>	<b>vp:ara</b>
Bengali	<b>ben</b>	<b>vp:ben</b>
Chinese, Japanese, and Korean	<b>cjk</b>	<b>vp:cjk</b>
Cyrillic	<b>cyr</b>	<b>vp:cyr</b>
Devanagari	<b>dev</b>	<b>vp:dev</b>

Greek	<b>gre</b>	<b>vp:gre</b>
Hebrew	<b>hbr</b>	<b>vp:hbr</b>
Tamil	<b>tam</b>	<b>vp:tam</b>
Thai	<b>tha</b>	<b>vp:tha</b>

To enter one of the searches above to retrieve all records that contain a specified script, use the command line in the Search WorldCat window (**Cataloging > Search > WorldCat**).

**Note:** If a search for a particular script alone retrieves too many WorldCat records (limit 1,500 records), you must limit the search and try again. (See more about how the client displays WorldCat search results in Cataloging, Search WorldCat, "Use WorldCat search results.")

### **Examples:**

**vp:ara/1991-2** (search for Arabic script records limited to those published in 1991 and 1992)

**vp:ara and la:per** (search for Arabic script records limited to those describing Persian language items)

See more about word and phrase searching and search methods in general in Cataloging/Search WorldCat:

- "Search WorldCat interactively"
- "Keyword, numeric, and derived search syntax" (derived search is unavailable for non-Latin script data)
- "Browse WorldCat"
- "Customize WorldCat search and browse interfaces"
- "Enter WorldCat searches for batch processing"

### **Sort order of search results**

You can select how the results of non-Latin script WorldCat searches are sorted:

- Alphabetically by the Latin script data  
Or
- In Unicode order by the non-Latin script data

To check or change the option for sort order for WorldCat search results:

	Action
1	On the Tools menu, click <b>Options</b> (or press <Alt><T><O>), and then click the <b>International tab</b> .
2	Click the <b>Primary Sort by Latin Script</b> check box to select or clear the option to sort search results in alphabetical order by the Latin script data. <b>Default:</b> Check box selected. Search results sort alphabetically by Latin script data. <b>Result:</b> <ul style="list-style-type: none"> <li>• If you clear the check box, search results are sorted in Unicode order by the non-Latin-script data.</li> <li>• The sort order selected also determines the sort order of local bibliographic save file and local constant data search results.</li> <li>• Tamil Unicode 4.0 codes are not in collating order. The default, alphabetical sorting by Latin script, is recommended if romanized (Latin-equivalent) data is included in the record with Tamil script data.</li> </ul>
3	When finished, click <b>Close</b> , or press <Enter> to apply the settings and close the Options window. Or Click <b>Apply</b> to apply the settings without closing the window.

**Display non-Latin scripts in records and lists**

In records that have paired non-Latin and Latin-equivalent script fields, the non-Latin script field appears first in the pair.

Arabic and Hebrew script data, by default, displays (and prints) right to left (**View > Align Right**). See "Catalog using Arabic script" or "Catalog using Hebrew script" for information about using Unicode formatting control characters to ensure correct display of bidirectional data in Arabic and Hebrew script records (use the right-click menu).

**Create records**

Use as many supported non-Latin scripts as you need anywhere in a record, including within the same field.

**Workforms**

For creating bibliographic records and/or constant data with non-Latin scripts using workforms (**Cataloging > Create > Single Record > [format]** or **Constant Data**), you can set an option to display the workforms with paired fields:

	Action
1	On the Tools menu, click <b>Options</b> (or press <Alt><T><O>), and then click the <b>International tab</b> .

	Action
2	<p>Click the <b>Include paired fields in workforms for multi-script data check box</b>.</p> <p><b>Default:</b> Check box cleared. Workforms contain single fields.</p> <p><b>Results:</b></p> <ul style="list-style-type: none"> <li>• The workform opens with paired fields 1XX, 245, 246, 250, 260, 300, 4XX, 5XX, 6XX, 8XX for non-Latin script entry (X = any valid tag number).</li> <li>• Each pair has identical tags.</li> <li>• The first of a paired field is for non-Latin script data.</li> </ul>
3	<p>When finished, click <b>Close</b>, or press &lt;Enter&gt; to apply the settings and close the Options window.</p> <p>Or</p> <p>Click <b>Apply</b> to apply the settings without closing the window.</p>

**Setting the option for paired fields is not required.** You can enter non-Latin script data only or romanized (Latin-script equivalent) data only. Or create your own paired fields and enter both:

	Action
1	In a record, add a field and enter the same tag as the corresponding existing field.
2	Enter the non-Latin script data in the first field of the pair, and optionally, enter the romanized data in the second field.
3	<p>To link the fields, on the Edit menu, click <b>Linking Fields &gt; Link Field</b>, or press &lt;Alt&gt;&lt;E&gt;&lt;K&gt;&lt;L&gt;.</p> <p>Or</p> <p>Let the client link them automatically when you validate, reformat, save, or take a final OCLC action on the record.</p>

**Caution:** If Latin script and non-Latin script parallel fields are not linked, display of records downloaded to your local system may be affected. You can set an option to get a warning before the client exports records with parallel unlinked non-Latin script fields in **Tools > Options > Export**.

### Derive records

When you use **Edit > Derive > New Master Record**, **New Institution Record**, or **New Constant Data** to create records from existing records that have linked fields for non-Latin scripts, the client transfers the linked fields as pairs for each field selected in **Tools > Options > Derive Record**.

**Note:** Although the 066 field cannot be transferred, the client adds the field automatically to indicate the presence and type of non-Latin script when you validate, reformat, save, or take a final OCLC action on the record.

See also general information about how to create bibliographic records in Cataloging/Create Bibliographic Records, "About creating bibliographic records"

### Edit records

#### Editing functions supported

- Find and Replace (**Edit** menu). Enter non-Latin script in both the **Find What** box and the **Replace With** boxes of the Find/Replace window.
- Cut, copy, and paste (**Edit** menu). Cut, copy, and paste non-Latin script data.
- When you move one of a paired field, the other field moves automatically (**Edit > Move Field > Up** or **Down**).
- Validate records and characters (**Edit** menu)
  - Validates non-Latin script characters against supported MARC-8 character sets, as well as validating MARC structure and tags.

Also validates Bengali, Devanagari, Thai, and Tamil characters coded in Unicode 4.0 (not covered in MARC-8 character sets).

If the client finds an invalid character, an error message lists the tag and position of the character to help you find it, along with other errors found in the record.

The error message may give up to three positions per field for invalid characters.

After correcting characters, you may want to validate the record again.

—Automatically adds field 066 indicating the presence and type of non-Latin script.

- Validate characters only. Validate characters separately from record validation (available for supported MARC-8 character sets only; unavailable for Bengali, Devanagari, Tamil, and Thai characters). Use **Edit > MARC-8 Characters > Verify**.

**Note for CJK:** You can automatically convert invalid CJK characters to equivalent MARC-8 characters. Use **Edit > MARC-8 Characters > Convert to MARC-8 CJK**.

- Reformat (**Edit** menu). Rearranges fields in MARC tag order, including:
  - Displaying paired fields together, with the non-Latin script field on top.
  - Automatically adding field 066, if not already added, with text that indicates the presence and type of script.
- Text strings (**Tools** menu). Use non-Latin scripts to create or edit text strings. See more about text strings in Basics, Set Options and Customize, "Create custom text strings."

#### Controlling headings unsupported

- Controlled headings in bibliographic records are linked directly to the controlling authority record so that if the authority record changes, the heading is automatically updated in the bibliographic record also.
- For non-Latin script records, however, only a heading in the Latin script field can be controlled (and subsequently updated if the authority record changes). If the

controlled heading is updated, you may need to update the corresponding non-Latin script field manually to match the update.

### Verify MARC-8 characters

To check the validity of characters separately from the validate records function (**Edit > Validate** or <Shift><F5>):

	Action
1	<p>On the Edit menu, click <b>MARC-8 Characters &gt; Verify</b>, or press &lt;Alt&gt;&lt;E&gt;&lt;8&gt;&lt;V&gt;.</p> <p><b>Results:</b></p> <ul style="list-style-type: none"> <li>The client changes the color of invalid characters to red by default (or a color you specify in <b>Tools &gt; Options &gt; Record Display</b>).</li> <li>If no invalid characters are present, you get a message that verification is completed. Click <b>OK</b> or press &lt;Enter&gt; to close.</li> </ul> <p><b>Note:</b> MARC-8 verification is inappropriate for Bengali, Devanagari, Tamil, and Thai characters, which are not covered in MARC-8 character sets. Bengali, Devanagari, Tamil, and Thai characters are verified when you validate records.</p> <p><b>Tip:</b> If the client identifies invalid CJK characters, you can use an automatic converter to convert them to equivalent MARC-8-compliant characters. (See "Catalog using Chinese, Japanese, and Korean (CJK) scripts.")</p>
2	<p>To remove invalid character display (display all text in the default text color or the color you selected in <b>Tools &gt; Options &gt; Record Display</b>), click <b>MARC-8 Characters &gt; Clear</b>, or press &lt;Alt&gt;&lt;E&gt;&lt;8&gt;&lt;C&gt;.</p> <p>See <b>Tools &gt; Options &gt; Record Display</b> for color options. Default color for invalid characters: <b>red</b>.</p>

**Tip:** If non-Latin characters that you need (other than Bengali, Devanagari, Tamil, and Thai) are not in any MARC-8 approved character sets for MARC 21 cataloging:

- Enter the character in the record, export the record to your local system using Unicode export format, and then remove the character before processing the record in WorldCat.

Or

- Enter the name of the character within square brackets, using the Unicode standard if available, (for example, enter **[schwa]**), or for CJK characters, enter the reading of the character (for example, enter **[yin]**).

For reference, see, for example, the Unicode charts Web page, which has a name index, at [www.unicode.org/charts/](http://www.unicode.org/charts/).

### Link/unlink paired non-Latin/Latin script fields

The client automatically links two non-Latin script/Latin script fields that have the same tag number when you validate, reformat, save, or take an action on the

record. The client always treats a non-Latin script field as the first of a corresponding pair.

You can link or unlink two non-Latin/Latin script fields with the same tag number:

Action
<p><b>To link fields:</b></p> <p>Place the cursor in the first field of a set of paired fields (the non-Latin data field), and on the Edit menu, click <b>Linking Fields &gt; Link Fields</b>, or press &lt;Alt&gt;&lt;E&gt;&lt;K&gt;&lt;L&gt; to link all paired fields.</p> <p>Or</p> <p>Right-click, and on the popup shortcut menu, click <b>Link Fields</b> to link the two fields where the cursor is located.</p> <p><b>To unlink fields:</b></p> <p>Click <b>Linking Fields &gt; Unlink Fields</b>, or press &lt;Alt&gt;&lt;E&gt;&lt;K&gt;&lt;U&gt; (unlinks all linked fields).</p> <p>Or</p> <p>Right-click, and on the popup shortcut menu, click <b>Unlink Fields</b> (unlinks the pair of fields where the cursor is located).</p>

**When you link fields:**

- The client uses a bracket to display linked fields, as in the following example showing part of a CJK record:

110	2	国立国会図書館 (Japan)
110	2	Kokuritsu Kokkai Toshokan (Japan)
245	1 0	国立国会図書館所蔵中国語・朝鮮語雑誌新聞目録 : #b 平成 3 年末現在 / #c [編集国立国会図書館専門資料部].
245	1 0	Kokuritsu Kokkai Toshokan shozō Chūgokugo, Chōsen-go zasshi shinbun mokuroku : #b Heisei 3-nenmatsu genzai / #c [henshū Kokuritsu Kokkai Toshokan Senmon Shiryōbu].
260		東京 : #b 国立国会図書館, #c 平成 5 [1993]
260		Tōkyō : #b Kokuritsu Kokkai Toshokan, #c Heisei 5 [1993]

- Printouts of records retain the brackets to indicate linked fields.
- If you modify the tag of one of the linked fields, the tag for the other field changes, too.
- If the cursor is in a linked field when you add a new field, the new field is added above or below the set of linked fields. Linked fields cannot be separated.
- Moving a linked field moves the set of linked fields.
- If you delete one field in a linked field set, the client keeps the other field and removes the link indicator (bracket).

**Caution:** If Latin script and non-Latin script parallel fields are not linked, display of the non-Latin script in records downloaded to your local system may be affected. You can set an option to get a warning before the client exports records with unlinked non-Latin script fields in **Tools > Options > Export**.

**Caution:** If Latin script and non-Latin script parallel fields are not linked, display of records downloaded to your local system may be affected.

### **Align Arabic and Hebrew data**

By default, the client displays (and prints) Arabic or Hebrew script data in Arabic, Persian, and Hebrew records aligned to the right. Toggle alignment for these scripts using **View > Align Right**. See more about aligning Arabic and Hebrew script data in these topics: "Arabic cataloging" or "Hebrew cataloging."

### **Use Unicode formatting characters for bidirectional Arabic and Hebrew data**

Valid left-to-right character strings (multiple digit numbers and punctuation) appear mixed in with right-to-left script data in Arabic, Persian, and Hebrew records. To ensure that this bidirectional data displays correctly, use Unicode formatting control characters.

The formatting control characters distinguish how to display mixed left-to-right and right-to-left data in an Arabic or Hebrew field. To insert a control character, right-click in a field, and on the pop-up menu click **Insert Unicode Control Character**. Then click a character.

For details, see the *Bidirectional Algorithm* report on the Unicode Web site at: <http://unicode.org/reports/tr9>.

See also "Cataloging using Arabic script" or "Cataloging using Hebrew script" in this booklet.

### **Use CJK E-Dictionary**

Use the CJK E-Dictionary (electronic dictionary) on the Tools menu to search or browse for information about:

- A single CJK character
- A group of related characters
- Homophones matching a phonetic input code
- A large set of characters in sequence by East Asian Character Code (EACC) or by Unicode value

The CJK E-Dictionary provides details on all CJK characters supported in Connexion client (see separate section, "Use CJK E-Dictionary").

### **Transliterate Arabic**

For existing Arabic records that contain only romanized data (Latin-script-equivalent representation of the Arabic script), the client provides two ways to automatically convert and add the equivalent romanized data (see "Catalog using Arabic scripts").

### Use constant data

- Use non-Latin scripts in workforms to create bibliographic constant data records (**Cataloging > Create > Constant Data**) or derive new constant data records from existing records (**Edit > Derive > New Constant Data**) in the same way as described above for creating and deriving bibliographic records.
  - For constant data in the local file only, you can use non-Latin scripts to name constant data records and search for them by non-Latin script name.
  - Also for constant data records in the local file only, you can use non-Latin scripts to specify My Status for records (**Action > Set Status**) and search for non-Latin script My Statuses. (My Status is an optional free-text status that you add to records to help distinguish them.)
  - You can change the sort order for results of searching the local constant data file to Unicode sorting by non-Latin script (default sort order: alphabetical sorting by Latin script). The same setting also determines sort order for WorldCat search results and local save file search results.
  - When you apply constant data containing non-Latin script, the client:
    - Replaces a non-repeatable field in the bibliographic record with the paired non-Latin script 880 and equivalent romanized field from the constant data.
- Or
- Adds the paired fields from the constant data below existing repeatable fields.

See general procedures for creating, applying, finding, and using bibliographic constant data in Cataloging/Use Bibliographic Constant Data.

### Save records and search save files

- Using the Connexion client only:
    - Save and display non-Latin scripts in records for editing in the online or local save file.
    - The non-Latin data also displays in local save file lists but does not display in online save file lists.
- Note:** If you open an online save file record containing non-Latin script in the browser, the non-Latin script data displays in 880 fields at the end of the record. The record opens in display mode only. You cannot edit or take actions on it.
- You can use non-Latin scripts to specify My Status for records in the online or local save file (**Action > Set Status**) and search for non-Latin script My Statuses. (My Status is an optional free-text status that you add to records to help distinguish them.)
  - **Local file indexes for non-Latin script data.** Use non-Latin scripts to search the following indexes for records in the local save file:
    - Name**  
Index includes non-Latin script fields associated with fields 100, 111, 130, 700, 710, 711, 730 (all subfields).

—**Title**

Index includes non-Latin script fields associated with fields/subfields:

- 245 a b f g k n p
- 246 a b f g n p

—**My Status**

- Change the sort order for results of searching the local save file to Unicode sorting by non-Latin script (**default** sort order: alphabetical by Latin script). The same setting also determines sort order for WorldCat search results and local constant data search results.
- When you save a record online or locally, the client automatically adds field 066 indicating the presence and type of non-Latin script.

See general procedures for saving, finding, and using bibliographic save file records in Cataloging/Save Bibliographic Records.

**Report errors in non-Latin script records**

You can use non-Latin script text in the message box of the Report Error window to report errors in non-Latin script records (**Action > Report Error**).

**Export records**

**Select options for exported data**

Select the type and location of Latin script versus non-Latin script data in exported records:

	<b>Action</b>
1	On the Tools menu, click <b>Options</b> (or press <Alt><T><O>), and then click the <b>International tab</b> .
2	Under <b>Export</b> , click one of the following check boxes: <ul style="list-style-type: none"> <li>• Include all data, with other scripts in 880 fields (<b>default</b>)</li> <li>• Include all data, with Latin script in 880 fields</li> <li>• Include Latin script only (deletes field 066 in record)</li> <li>• Include other scripts only</li> </ul>
3	When finished, click <b>Close</b> , or press <Enter> to apply the settings and close the Options window. Or Click <b>Apply</b> to apply the settings without closing the window.

**Select character set**

- Select either MARC-8 (**default**) or UTF-8 Unicode to export records with Arabic, CJK, Cyrillic, Greek, or Hebrew scripts. These MARC-8 character sets are subsets of UTF-8 characters, so selecting either character set covers them.
- Use only UTF-8 Unicode to export records with Bengali, Devanagari, Tamil, and Thai scripts. These scripts are not included in MARC-8.
- If any non-MARC-8 scripts are exported in MARC-8 data format, the non-MARC-8 characters are saved in Numeric Character Reference (NCR) format (see [http://en.wikipedia.org/wiki/Numeric\\_Character\\_Reference](http://en.wikipedia.org/wiki/Numeric_Character_Reference) for more information).

To select a character set for exporting records:

	Action
1	On the Tools menu, click <b>Options</b> (or press <Alt><T><O>), and then click the <b>Export tab</b> .
2	Click <b>Record Characteristics</b> .
3	Under <b>Bibliographic Records</b> , select one of the following from the <b>Character Set</b> list: <ul style="list-style-type: none"> <li>• <b>MARC-8 (default)</b></li> <li>• <b>UTF-8 Unicode</b></li> </ul>
4	Click <b>OK</b> , or press <Enter> to save your settings and close the window, or click <b>Cancel</b> to cancel changes.  You are returned to the Export page in the Options window.  When finished, click <b>Close</b> to close the Options window.

### Select fields to delete from exported records

	Action
1	In the <b>Export Options</b> tab ( <b>Tools &gt; Options &gt; Export</b> ), click <b>Field Export Options</b> .
2	Under <b>Fields to Delete</b> , in the <b>Bibliographic Records</b> text box, enter tag numbers for fields you want to delete in exported bibliographic records.  Separate numbers by a comma and a space, or use a hyphen to show a range.  <b>Example:</b> 920, 938-999
3	Repeat step 2 in the <b>Authority Records</b> text box if needed for exporting authority records.
4	When finished, click <b>OK</b> , or press <Enter> to save your settings and close the window, or click <b>Cancel</b> to cancel changes.  You are returned to the Export page in the Options window.  When ready, click <b>Close</b> to close the Options window.

### Set warning before exporting records with parallel unlinked non-Latin script fields

If Latin script and non-Latin script parallel fields are not linked, display of the non-Latin script in records downloaded to your local system may be affected.

Set an option to get a warning when exporting before the client exports records with unlinked non-Latin script fields in **Tools > Options > Export**.

See general instructions for exporting records in **Cataloging/Export Bibliographic Records**.

**Import records**

- Select either MARC-8 (**default**) or UTF-8 Unicode to import records with Arabic, CJK, Cyrillic, Greek, or Hebrew scripts. These MARC-8 character sets are subsets of UTF-8 characters, so selecting either character set covers them.
- Use only UTF-8 Unicode to import records with Bengali, Devanagari, Tamil, and Thai scripts. These scripts are not included in MARC-8.
- If any non-MARC-8 scripts are imported using the MARC-8 character set option, the non-MARC-8 characters are saved in Numeric Character Reference (NCR) format (see [http://en.wikipedia.org/wiki/Numeric\\_Character\\_Reference](http://en.wikipedia.org/wiki/Numeric_Character_Reference) for more information).

To select a character set for importing records:

	Action
1	On the File menu, click <b>Import Records</b> (or press <Alt><F><I>).
2	In the Import Records window, specify the import file and destination, and then click <b>Record Characteristics</b> .
3	Under <b>Bibliographic Records</b> , select one of the following from the <b>Character Sets</b> list: <ul style="list-style-type: none"> <li>• <b>MARC-8 (default)</b></li> <li>• <b>UTF-8 Unicode</b></li> </ul> Note: Always select <b>UTF-8 Unicode</b> for Bengali, Devanagari, Tamil, and Thai scripts.
4	Click <b>OK</b> or press <Enter> to save your settings and close the window, or click <b>Cancel</b> to cancel changes. You are returned to the Import Records window.

See general procedures in Cataloging/Import Bibliographic Records.

**Using non-Latin script data in macros**

You can incorporate non-Latin script data in Connexion client macros using SetFieldUnicode, SetFieldLineUnicode, and GetListCellDataUnicode. Data is converted to Numeric Character Reference (NCR) format.

See detailed descriptions of these macros in Basics/Use Macros, "Connexion client macro commands: Edit records."

### 3 Add non-Latin script variant name headings in authority records

#### About using non-Latin scripts for variant name headings in LC authority file records

The Library of Congress and other major authority record exchange partners—British Library, National Library of Medicine, and OCLC, in consultation with the Library and Archives Canada—are implementing the use of non-Latin scripts in records for Name Authority Cooperative Program (NACO) contribution/distribution processes.

When completed (not before April 2008), you will be able to use non-Latin script variant forms of name headings in:

- Fields 4XX and 7XX
- Various note fields (for example, 67X)

NACO participants can add non-Latin script data to master LC name authority records. Non-NACO catalogers can add non-Latin scripts and export records for local use. Only MARC-8 character sets for the following scripts will be supported:

- Arabic (including the Persian language)
- Chinese
- Cyrillic
- Greek
- Hebrew (including Yiddish)
- Japanese
- Korean

In addition to creating, editing, and saving non-Latin script data in authority records, searching, browsing, displaying, saving, applying constant data, and exporting is available for authority records containing non-Latin script records.

Versions 2.10 and higher of the client will support using non-Latin scripts for variant heading and notes fields as soon as this functionality becomes available. **The actual date of availability will be no earlier than April 2008.** OCLC will announce more details beforehand.

#### Details

- The Latin script or romanized form of a heading in field 1XX will continue to be the authorized heading.
- The LC authority file will not have paired records for Latin script and non-Latin script forms of name headings for the same entity.
- NACO contributors will follow MARC 21's "Model B" for multiscript records. **Model B provides for unlinked non-Latin script fields**, such as authority record 4XX fields, that have the same MARC tags used for Latin script data.
- Using Model B for authorities is a departure from the current bibliographic record practice of many Anglo-American Cataloging libraries, where non-Latin characters are exported as linked 880 fields (Alternate Graphic Representation) using MARC 21's "Model A" for multiscript records.

- Although Connexion client supports Bengali, Devanagari, Tamil, and Thai for use in bibliographic records, character sets for these scripts will not be supported for authority records.

See more information on the Library of Congress Web site at: [http://www.loc.gov/catdir/cpsa/nonroman\\_announce.pdf](http://www.loc.gov/catdir/cpsa/nonroman_announce.pdf).

### Character sets supported

MARC-8 character sets for non-Latin scripts to be available for references in authority records are listed in *MARC 21 Specifications for Record Structure, Character Sets, and Exchange Media, Code Tables* (see <http://www.loc.gov/marc/specifications/specchartables.html>).

The following supported character sets are subsets of UTF-8 Unicode that are approved for use in MARC 21 cataloging:

- Basic Arabic = 33 (hex)
- Extended Arabic = 34 (hex)
- Chinese, Japanese, Korean (EACC) = 31 (hex)
- Basic Cyrillic = 4E (hex)
- Extended Cyrillic = 51 (hex)
- Basic Greek = 53 (hex)
- Basic Hebrew = 32 (hex)

### Existing non-Latin script support in Connexion client for bibliographic records: What applies to authority records?

Much of the client Help information about using non-Latin scripts for bibliographic records also applies to using supported scripts for authority records; for example, see more about input methods for languages that use non-Latin scripts in "Connexion client international cataloging."

### Functions that are the same for authority records

Functions that apply to both bibliographic and authority records include:

- A single record and a single field in a record can have multiple non-Latin scripts, as in bibliographic records.
- Data alignment for displaying and printing Arabic and Hebrew script data (**View > Align Right**).
- MARC-8 character verification (verification separate from record validation) (**Edit > MARC-8 Characters > Verify** [or **Clear**]).
- Conversion of invalid CJK to MARC-8 (**Edit > MARC-8 Characters > Convert to MARC-8 CJK**).
- Manual transliteration of romanized data to Arabic or Persian using **Edit > Transliterate > Arabic** [or **Persian**].  
**Note:** Automatic transliteration (set by an option) is **not** available for name authority records.

- Unicode formatting control characters to support correct display of bidirectional data in Arabic and Hebrew script records (right-click in a field and click **Insert Unicode Control Character** to choose a control character).  
**Note:** Formatting control characters are **not** automatically added to Arabic or Hebrew data. You must add them manually. They are available **only** from the right-click menu.
- CJK E-Dictionary, which helps with character selection by providing comprehensive information about CJK characters supported in the client (**Tools > CJK E-Dictionary**).
- Use non-Latin script search terms in all LC authority file search indexes to retrieve a particular record or set of records.
- If you want to retrieve all records or see sample records containing a particular script, use the "character sets present" search index (label **vp:**) with the assigned code for a script:

Script	Code for script	Enter search as ...
Arabic	<b>ara</b>	<b>vp:ara</b>
Chinese, Japanese, and Korean	<b>cyj</b>	<b>vp:cyj</b>
Cyrillic	<b>cyr</b>	<b>vp:cyr</b>
Greek	<b>gre</b>	<b>vp:gre</b>
Hebrew	<b>hbr</b>	<b>vp:hbr</b>

To enter one of the searches above to retrieve all records that contain a specified script, use the command line in the LC Names and Subjects Search window (**Authorities > Search > LC Names and Subjects**).

See more about how the client displays LC authority file search results in Authorities, Search Authority Files, "Use LC authority file search and browse results."

### Functions that are different for authority records

The following features function differently for non-Latin script data in authority records:

- **Character set identifier.** No script identifier will appear in authority master records (unlike bibliographic records which have identifiers in field 066 ꞵc).
- **Validation.** Validation of authority records is, of course, different from validation of bibliographic records. Specifically, validation of authority records that contain non-Latin script data:
  - Is limited to name authority records (no sh/sj)
  - Is limited to the following heading fields: 400, 410, 411, 430, 451, 700, 710, 711, 730, 751
  - Is limited to some 6XX notes fields (to be determined)
  - Checks for display from left to right or right to left based on the Unicode range of the first character after the first subfield code

- Allows adding or replacing authority records only for MARC-8 character sets.
- Allows exporting authority records with non-MARC-8 data added locally only if validation is set to **None** in **Tools > Options > General** (click **Validation Level Options**).

### Functions that are not available for authority records

The following existing functionality for bibliographic records **is not available** for authority records:

- Supported scripts: For bibliographic records, the client also supports character sets for Bengali, Devanagari, Tamil, and Thai. However, these scripts are not in the MARC-8 repertoire of UTF-8 character sets, putting them out of the scope of this implementation of non-Latin scripts for authority records. They will not be supported for authority records.
- Export and import character set option - always use the **default** MARC-8 character set for authority records.
- Export options for data fields and options for transliteration in **Tools > Options > International** are unavailable.
- Automatic Arabic and Persian transliteration is unavailable (note that manual transliteration is available).
- Field linking/unlinking - non-Latin script fields in authority records are always unlinked.

## 4 Catalog using Arabic scripts

### About using Arabic scripts

**Cataloging.** Use Arabic script data for cataloging items in languages that use the Arabic script (for example, besides Arabic, Persian, Urdu, and Azerbaijani). Use Arabic script data the same way you use other non-Latin script data in the client. See "Use non-Latin scripts for cataloging" for details on using non-Latin scripts and records. See client Help or other Connexion Client System Guides on the OCLC Web site for general procedures describing how to:

- Search WorldCat
- Create records
- Edit records
- Use constant data
- Save records
- Take OCLC actions
- Export records
- Import records

**Authorities.** Arabic scripts can also be used to add variant name headings to authority records. See section above for more information.

### Tools for using non-Latin scripts

#### Specific tools to help with Arabic script cataloging

- Automatically transliterate existing romanized data (Latin script equivalent) into Arabic script data for Arabic and Persian records:
  - Use **Edit > Transliterate > Arabic** [or **Persian**] to transliterate selected data in a record.
  - Set an option (for bibliographic records only) in **Tools > Options > International** to auto-transliterate Arabic and/or Persian WorldCat records when you download them interactively.
- Toggle alignment for Arabic or Hebrew script data right-to-left or left-to-right using **View > Align Right** (default: right-to-left)
- Use Unicode formatting characters to control correct display of bidirectional data in Arabic and Hebrew records.
- See procedures below for using these Arabic-specific tools.

#### Other tools to help with non-Latin script cataloging in general

- MARC-8 character verification (**Edit > MARC-8 Characters > Verify**) - verify characters separately from record validation.
- Link/unlink fields (**Edit > Linking Fields > Link** [or **Unlink**]) - visually link non-Latin script data fields with equivalent Latin script (romanized) data fields (bibliographic records only).

- Export options for data fields (**Tools > Options > International**) - determine (for bibliographic records only):
  - Whether to export both Latin-script-equivalent (romanized) data and non-Latin script data or only one or only the other
  - Position of data if you export both Latin and non-Latin script data
  - Sort order
- Export and import data using UTF-8 Unicode or MARC-8 character sets. The UTF-8 Unicode option allows you to work with non-MARC-8 characters in the client for your local records (settings for export are in **Tools > Options > Export**, click **Record Characteristics**, and settings for import are in **File > Import Records**, click **Record Characteristics**).
- See "Use non-Latin scripts for cataloging" for more specific procedures for working with these tools.

### Arabic script entry and character sets

#### Script entry methods

- If your system default language is not Arabic, you can install the Arabic language (various forms) in Windows. When you install Arabic, Windows provides an input keyboard for entering Arabic script. See "Connexion client international cataloging," "Input methods for languages that use non-Latin scripts."
- OCLC provides an alternative Arabic script keyboard developed for RLIN21 cataloging software. You can download the Arabic keyboard from the OCLC Web site. For instructions, see *Getting Started* at <http://www.oclc.org/support/documentation/connexion/client/gettingstarted/gettingstarted>.

See also, *RLIN21 Keyboards* for graphic illustrations of all keyboards, including Arabic, at (<http://www.oclc.org/support/documentation/connexion/client/gettingstarted/gettingstarted/rlin21keyboards.pdf>.)

RLIN21 keyboards include characters specific to each script (covering multiple languages that use Arabic script), whereas Microsoft keyboards include script characters specific to a single language.

#### Character sets supported

The client supports the basic and extended Arabic character sets defined in *MARC 21 Specifications for Record Structure, Character Sets, and Exchange Media*.

These character sets are available on the Library of Congress Web site at:

< <http://lcweb2.loc.gov/cocoon/codetables/33.html> > (Basic)

< <http://lcweb2.loc.gov/cocoon/codetables/34.html> > (Extended)

- 33(hex) [ASCII graphic: 3] = Basic Arabic
- 34(hex) [ASCII graphic: 4] = Extended Arabic

### Script identifier in records

The client adds the following data to ꜥc of field 066 in Arabic records to indicate the presence of Arabic characters:

- (3 (Basic Arabic)
- (4 (Extended Arabic)

### Romanized data

See the *ALA-LC Romanization Tables* for Arabic and for Persian on the Library of Congress Web site at:

<http://www.loc.gov/catdir/cpsa/romanization/arabic.pdf> (Arabic)

<http://www.loc.gov/catdir/cpsa/romanization/persian.pdf> (Persian)

### Transliterate romanized data in Arabic or Persian records into Arabic script

The client provides two ways to automatically transliterate existing romanized data into Arabic script data:

- Use **Edit > Transliterate > Arabic [or Persian]** to transliterate romanized data in selected fields of a displayed record.
- Set an option (for bibliographic records only) in **Tools > Options > International tab** to auto-transliterate romanized data in all Arabic records retrieved interactively from WorldCat (records with language code **ara** or **per** and no field 066). Also select the fields to auto-transliterate.


### Transliterate selected fields in a record

	Action
1	Display a bibliographic, authority, or constant data record containing romanized data that describes Arabic language materials.
2	Place the cursor in the field containing romanized data that you want to transliterate. Or Select multiple fields containing romanized data. If you select parts of fields, the client transliterates the entire field(s).
3	Click <b>Edit &gt; Transliterate &gt; Arabic [or Persian]</b> , or press <Alt><E><T><A> or <Alt><E><T><P>, or right-click and on the pop-up menu, click <b>Transliterate &gt; Arabic [or Persian]</b> .  See "Results of transliteration" below.

**Note:** Although you can transliterate into Arabic while working offline (you do not need to be logged on to the OCLC system), your workstation must have an Internet connection.

**Auto-transliterate WorldCat records retrieved interactively (for bibliographic records only)**


Alternatively, select an option to auto-transliterate romanized data in all WorldCat records you retrieve interactively when the records have the language code **ARA** but no field 066:

	Action
1	Click <b>Tools &gt; Options</b> (or press <Alt><T><O>), and then click <b>International</b> .
2	Click to select the <b>Auto-transliterate Arabic records</b> check box and/or the <b>Auto-transliterate Persian records</b> check box.  <b>Note:</b> This option works for records that have the language code <b>ara</b> or <b>per</b> but no field 066.
3	<b>Optional.</b> Select fields to auto-transliterate: 1. In the International window, select the <b>Auto-transliterate Arabic fields</b> check box and/or the <b>Auto-transliterate Persian fields</b> check box. 2. Click <b>Choose Fields</b> to open the Choose Fields to Auto-Transliterate window. 3. Click to select or clear check boxes next to fields 1XX through 8XX ( <b>X</b> = any valid tag number). <b>Default:</b> Fields 1XX through 8XX are selected. 4. Click <b>OK</b> to save your settings or <b>Cancel</b> to cancel changes. You are returned to the <b>International</b> window.
4	When finished, click <b>Close</b> , or press <Enter> to apply the settings and close the Options window. Or Click <b>Apply</b> to apply the settings without closing the window. Or Click <b>Cancel</b> to cancel changes.  See "Results of transliteration" below.  <b>Note:</b> When you retrieve and display WorldCat records, the client marks any auto-transliterated fields with the symbol 

**Results of transliteration and auto-transliteration**

The client:

- Transliterates the romanized data word by word, independently of context.  
  
**Note:** If context other than that of letters within a word is a factor in the appearance of the Arabic text, you may need to edit the Arabic transliteration. See also the **caution** below.
- Creates an identical field with the same tag number (for example, two 245 tags) to contain the transliterated Arabic script.
- Places the Arabic script field above the associated romanized data field.
- Links the pair of associated fields with a bracket.

- If auto-transliterated (option for bibliographic records only, selected in **Tools > Options > International**), marks transliterated fields with the symbol 

**Caution:** Transliteration handles the following characters incorrectly. Revise the characters manually.

- The final character taa' marbuta preceded by hamza transliterates incorrectly as haa'.
- When 'alif maksura is followed by a period, the transliteration omits 'alif maksura.
- 'Alif laam followed by 'alif madda transliterates incorrectly as 'alif laam 'alif.
- Hyphens are incorrectly deleted in transliterated text.
- When laam kasra is followed by siin or jiim, the transliteration omits siin or jiim.
- Laam kasra followed by Haa' transliterates incorrectly as haa'.
- When two laams are followed by capital A (where the first laam is a preposition), the transliteration omits 'alif hamza. However, laam followed by lowercase "a" transliterates correctly as laam 'alif.
- When laam hyphen is followed by damma, the transliteration omits 'alif hamza.

**Basis of transliteration** The client transliterates romanized data based on the rules for Arabic given in *ALA-LC Romanization Tables* on the Library of Congress Web site at: < <http://www.loc.gov/catdir/cpsd/romanization/arabic.pdf> >.

**Align Arabic or Hebrew script data for display and print** By default, the client displays (and prints) Arabic or Hebrew scripts with data aligned to the right. To toggle between displaying these scripts right-to-left or left-to-right:

Action
<p><b>Toggle alignment for all Arabic or Hebrew script data in the current record:</b> Click <b>View &gt; Align Right</b>, or press &lt;Alt&gt;&lt;V&gt;&lt;I&gt;. <b>Default:</b> Data aligns to the right for display and printing.</p> <p><b>Result:</b> The Align Right icon next to the command on the View menu is active (highlighted) if Align Right is selected. The icon is inactive (grayed out) if Align Right is cleared.</p> <p>Or</p> <p><b>Toggle data alignment in the current field:</b> Right-click a field, and on the pop-up menu, click <b>Right-to-Left Reading Order</b>.</p> <p><b>Result:</b> The client changes alignment of the Arabic or Hebrew script data only in the current field.</p>

**Use Unicode formatting characters to control bidirectional data** Enter Unicode formatting characters in Arabic, Persian, and Hebrew records to correctly display left-to-right multiple-digit numbers and punctuation, including brackets, hyphens, internal spaces, etc., within a field of right-to-left script data.

- **Export/import using UTF-8 Unicode character set.** Unicode formatting control characters are retained as is in Arabic, Persian, and Hebrew records exported or imported using the UTF-8 Unicode character set, along with other non-MARC-8 Unicode characters.
- **Export/import using MARC-8 character set.** The Unicode formatting characters are retained in Numeric Character Reference (NCR) format in records exported or imported using the MARC-8 character set, along with other non-MARC-8 characters.

	Action
1	Click to locate the cursor in the position where you want to insert a formatting control number.
2	Right-click in the field, and on the pop-up menu click <b>Insert Unicode Control Character</b> . Or Right-click to open the pop-up menu and then press the keystrokes shown in step 3.
3	Click one of the following characters (or press the keystroke shortcuts, shown in parentheses): <ul style="list-style-type: none"> <li>• <b>LRM Left-to-Right Mark</b> (&lt;Alt&gt;&lt;R&gt;&lt;L&gt;)</li> <li>• <b>RLM Right-to-Left Mark</b> (&lt;Alt&gt;&lt;R&gt;&lt;R&gt;)</li> <li>• <b>ZWJ Zero Width Joiner</b> (&lt;Alt&gt;&lt;R&gt;&lt;J&gt;)</li> <li>• <b>ZWNJ Zero Width Non-Joiner</b> (&lt;Alt&gt;&lt;R&gt;&lt;N&gt;)</li> <li>• <b>LRE Start of Left-to-Right Embedding</b> (&lt;Alt&gt;&lt;R&gt;&lt;S&gt;)</li> <li>• <b>RLE Start of Right-to-Left Embedding</b> (&lt;Alt&gt;&lt;R&gt;&lt;T&gt;)</li> <li>• <b>LRO Start of Left-to-Right Override</b> (&lt;Alt&gt;&lt;R&gt;&lt;A&gt;)</li> <li>• <b>RLO Start of Right-to-Left Override</b> (&lt;Alt&gt;&lt;R&gt;&lt;F&gt;)</li> <li>• <b>PDF Pop Directional Formatting</b> (&lt;Alt&gt;&lt;R&gt;&lt;P&gt;)</li> </ul> <p><b>Tip for one-step entry:</b> Create a text string using <b>Tools &gt; Text Strings</b>; click <b>Add</b> and enter one of the characters listed above using the right-click menu. Then use the Text Strings quick tool on the toolbar to enter the character.</p> <p>Or Assign the text string to a keystroke shortcut. Enter the character by pressing the keystroke.</p>

**Example:** To control the display of the data *742[1981 or 1982]* that you enter in field 260 Ꞇc, and that is preceded and to be followed by Arabic script data:

1. Click to locate the cursor in field 260 Ꞇc.
2. Right-click in the field, and in the pop-up menu click **Insert Unicode Control Character**. Then click **LRE Start of Left-to-Right Embedding**.
3. Enter the data string, **742[1981 or 1982]**, immediately following the character.
4. Without moving the cursor, right-click in the field again. In the pop-up menu click **Insert Unicode Control Character**. Then click **PDF Pop Directional Formatting**.

### More information:

- For details, see the *Bidirectional Algorithm* report on the Unicode Web site at: <http://unicode.org/reports/tr9/>
- See more about selecting a character set for exporting and importing bibliographic records in the client in Cataloging, Export or Import Bibliographic Records.

### Use Arabic definite article in Arabic script searches

Always include the Arabic definite article ا (‘alif laam) in all words in a keyword search.

### Indexing for Arabic script searches

#### Notes on searching:

- Use word or phrase search indexes and word or phrase browse indexes.
- Word searches find the data string you enter anywhere in the indexed field. Phrase searches find the data string starting with the first character in a field or subfield and including each character in exact order. Browsing scans an index for the closest match to the character string followed by any other data.
- If you use qualifiers to limit searches, enter them using Latin script.
- Do not use derived searching.
- Do not use truncation (asterisk (\*)) at the end of a search term). You can use browsing for automatic truncation (enter only as many characters as needed for a match without using an asterisk at the end).
- If you want to retrieve all Arabic script records or see sample records, use the "character sets present" WorldCat search index (label **vp:**) with the assigned code **ara**.

#### To find all Arabic script records:

Enter **vp:ara** in the command line search of the Search WorldCat window (**Cataloging > Search > WorldCat**).

**Note:** If a search for all Arabic script records alone retrieves too many WorldCat records (limit 1,500 records), you must limit the search and try again.

#### Examples:

**vp:ara/1991-2**

**vp:ara and la:per**

See general procedures for searching WorldCat in the Cataloging/Search WorldCat booklet or client Help.

#### Arabic character indexing specifics:

The following table shows Arabic characters grouped together and indexed the same as if they are the same character (the characters are "normalized").

Type any character of a group of normalized characters in a search and retrieve results for all characters in the group.

Images and names of characters indexed the same are in columns 3 and 4, opposite the character with which they are indexed.

Character	Character name	Other characters indexed the same	
		Character	Character name
ا	'alif	ا	double 'alif with hamza above
		آ	'alif with madda above
		أ	'alif with hamza above
		إ	'alif with hamza below
		آء	'alif wasla
		أٓ	'alif with wavy hamza above
		إٓ	'alif with wavy hamza below
ت	taa'	ت	taa' marbuta
		تٓ	taa' with ring
		تٓٓٓ	taa' with three dots above
ه	Haa'	ه	Haa' with hamza above
		هٓ	Haa' with two dots vertical above
		هٓٓٓ	Haa' with three dots above
د	daal	دٓ	daal with ring
		دٓٓ	daal with dot below
		دٓٓٓ	daal with dot below and small Taa'
		دٓٓٓٓ	daal with three dots above downwards
		دٓٓٓٓٓ	daal with four dots above

Character	Character name	Other characters indexed the same	
		Character	Character name
ر	raa'	ر	raa' with small v
		ر	raa' with ring
		ر	raa' with dot below
		ر	raa' with small v below
		ر	raa' with dot above and below
		ر	raa' with two dots above
		ر	raa' with four dots above
س	siin	س	siin with dot below and dot above
		س	siin with three dots below
ش	shiin	ش	shiin with three dots below and three dots above
		ش	shiin with dot below
ص	Saad	ص	Saad with two dots below
		ص	Saad with three dots above
ض	Daad	ض	Daad with dot below
غ	ghayn	غ	ghayn with dot below
ط	Taa'	ط	Taa' with three dots above
ع	ayn	ع	ayn with three dots above

Character	Character name	Other characters indexed the same	
		Character	Character name
ف	faa'	ف	dotless faa'
		ف.	faa' with dot moved below
		ف.	faa' with dot below
		ف.	faa' with three dots below
ق	qaaf	ق	qaaf with dot above
		ق	qaaf with three dots above
ك	kaaf	ك	swash kaaf
		ك	kaaf with ring
		ك	kaaf with dot above
		ك	kaaf with three dots below
گ	gaf	گ	gaf with ring
		گ	gaf with two dots below
		گ	gaf with three dots above
ل	laam	ل	laam with small v
		ل	laam with dot above
		ل	laam with three dots above
		ل	laam with three dots below



## 5 Catalog using Bengali script

### About using Bengali script

Use Bengali script data for cataloging items in languages that use the Bengali script (for example, Bangla and Assamese). Use Bengali script data the same way you use other non-Latin script data in the client. See "Use non-Latin scripts for cataloging" for details on using Bengali script and Bengali records. See client Help or other Connexion Client System Guides on the OCLC Web site for general procedures describing how to:

- Search WorldCat
- Create records
- Edit records
- Use constant data
- Save records
- Report errors in records
- Export records
- Import records

### Tools for using non-Latin scripts

The client provides the following general tools to help you catalog using non-Latin scripts:

- Link/unlink fields (**Edit > Linking Fields > Link [or Unlink]**) - visually link non-Latin script data fields with equivalent romanized data fields.
- Export options for data fields (**Tools > Options > International**) - determine:
  - Whether to export both Latin-script-equivalent (romanized) data and non-Latin script data or only one or the other
  - Position of data if both
  - Sort order
- **Caution:** MARC-8 character verification (**Edit > MARC-8 Characters > Verify**) is not appropriate for verifying Bengali characters. There is no MARC-8 character set for Bengali. Using this command for Bengali results in marking all Bengali characters as invalid. The OCLC system validates Bengali characters when you validate a record.
- See "Use non-Latin scripts for cataloging" for more specific procedures for working with these tools.

### Unicode export and import required for Bengali records

Because Bengali script is not included in MARC-8 character sets, you must export and import records in Unicode format (settings are in **Tools > Options > Export** and **File > Import Records/Options** button).

### About Unicode

**Definition.** Unicode is the universal character encoding scheme for written characters and text. It defines a consistent way of encoding multi-script text that enables the exchange of text data internationally.

Unicode provides for three encoding forms: a 32-bit form (UTF-32), a 16-bit form (UTF-16), and an 8-bit form (UTF-8, designed for use with ASCII-based systems).

### Unicode Standard, Version 4.0

- Contains 96,382 characters from scripts used for languages worldwide.
- Identical to International Standard ISO/IEC 10646:2003, *Information Technology Universal Multiple-Octet Coded Character Set (UCS) Architecture and Basic Multilingual Plane, Supplementary Planes*, known as the Universal Character Set (UCS)
- *The Unicode Standard*, Version 4.0. The Unicode Consortium. Addison-Wesley Developers Press, 2003.

### Bengali script entry and character set

#### Script entry method

If your system default language is not Bengali, you can install Bengali, and Windows provides an input keyboard for entering Bengali script.. See "Connexion client international cataloging/Input methods for languages that use non-Latin scripts."

#### Character set supported

Bengali characters are defined in Unicode 4.0 (coded in the range U+0981 to U+09FA).

**Caution:** In Windows XP, the font size used to display Bengali script is too small. You can increase the font size for viewing and editing these records in **Tools > Options > Fonts. Font size for Bengali script is not a problem in Windows Vista.**

#### Script identifier in records

The client adds the following data to field 066 Ꞥc in Bengali records to indicate the presence of Bengali characters:

- **Beng**

#### Romanized data

See the *ALA-LC Romanization Table* for Bengali on the Library of Congress Web site at <http://www.loc.gov/catdir/cpsd/romanization/bengali.pdf>.

### Indexing for Bengali script searches

#### Notes on searching:

- Use word or phrase search indexes and browse indexes.
- Word searches find the data string you enter anywhere in the indexed field. Phrase searches find the data string starting with the first character in a field or subfield and including each character in exact order. Browsing scans an index for the closest match to the character string followed by any other data.
- If you use qualifiers to limit a search, type them in Latin script.
- Do not use derived searching.

- You can truncate searches (asterisk (\*)) at the end of a search term) or use browsing for automatic truncation (enter only as many characters as needed for a match, without using an asterisk).
- If you want to retrieve all Bengali script records or see sample records, use the "character sets present" WorldCat search index (label **vp:**) with the assigned code **ben**.

### **To find all Bengali script records:**

Enter **vp:ben** in the command line search of the Search WorldCat window (**Cataloging > Search > WorldCat**).

**Note:** If a search for all Bengali script records alone retrieves too many WorldCat records (limit 1,500 records), you must limit the search and try again.

### **Examples:**

**vp:ben/1991-**

**vp:ben and mt:bks**

See general procedures and search techniques for searching WorldCat in the Cataloging, Search WorldCat booklet or client Help.

### **Bengali character indexing specifics:**

- Bengali signs are indexed as is (Candrabindu, Anusvara, Visarga, Nukta, and Avagraha).
- Independent vowels, dependent vowels, two-part dependent vowels, and generic or Bengali-specific character additions are all indexed as is.
- Consonants are indexed as is, if attached with Virama (Hasant); otherwise, they are indexed with the dependent vowel or consonant.
- Both Bengali and Latin numbers are indexed (either may appear in Bengali text).

### **Notes on sorting search results**

- Bengali syllables with candrabindu or anusvara (nasalization signs) precede terms without those syllables.
- Non-conjunct forms of a consonant precede conjunct forms.
- The default sort order for search results alphabetical sorting by Latin script is recommended if romanized (Latin-equivalent) data is included in the record. The sort order option is in **Tools > Options > International**.

## 6 Catalog using Chinese, Japanese, and Korean (CJK) scripts

### About using CJK scripts

**Cataloging.** Use CJK script data to catalog items in Chinese, Japanese, and Korean. Use CJK script data the same way you use other non-Latin script data in the client. See "Use non-Latin scripts for cataloging" for details on using non-Latin scripts and records. See client Help or other Connexion Client System Guides on the OCLC Web site for client procedures describing how to:

- Search WorldCat
- Create records
- Edit records
- Use constant data
- Save records
- Report errors in records
- Export records
- Import records

**Authorities.** CJK scripts can also be used to add variant name headings to authority records (see section above for more information).

### Tools for using non-Latin scripts

The client provides the following general tools to help you catalog using non-Latin scripts:

- MARC-8 character verification (**Edit > MARC-8 Characters > Verify**) - verify characters separately from record validation.
- Link/unlink fields (**Edit > Linking Fields > Link [or Unlink]**) - visually link non-Latin script data fields with equivalent romanized data fields (for bibliographic records only).
- Export options for data fields (**Tools > Options > International**) - determine (for bibliographic records only):
  - Whether to export both Latin-script-equivalent (romanized) data and non-Latin script data or only one or the other
  - Position of data if both
  - Sort order
- Export and import data in Unicode or MARC-8 format - Unicode option allows you to work with non-MARC-8 characters that are unsupported in the client and in the WorldCat database for your local records (**Tools > Options > Export** and **File > Import Records/Options** button).
- See "Use non-Latin scripts for cataloging" for more specific procedures for working with these tools.
- **Specifically for CJK:** The client provides a:
  - CJK E-Dictionary to help with entering CJK characters: (**Tools > CJK E-Dictionary**).

—Automatic converter for converting invalid characters to equivalent MARC-8 characters (**Edit > MARC-8 Characters > Convert to MARC-8 CJK**).

—Chinese name authority file (**Authorities > Search > Chinese Name Authority File**). Access to records is read-only, but you can copy and paste from records or print them. See Search the Chinese name authority file and Chinese name authority file indexes for details.

### CJK entry and character set

#### Script entry method

If your system default language is not the one you want to use for cataloging Chinese, Japanese, or Korean materials, you can install the languages you need. Windows provides the Input Method Editors (IMEs) appropriate for CJK character entry. See more about script entry methods in "Connexion client international cataloging/Input methods for languages that use non-Latin scripts."

#### Character sets supported

The client supports the following Chinese, Japanese, and Korean character set defined in *MARC 21 Specifications for Record Structure, Character Sets, and Exchange Media*. These character sets are available on the Library of Congress Web site at: <http://www.loc.gov/marc/specifications/specchareacc.html>.

- 31(hex) [ASCII graphic: 1] = Chinese, Japanese, Korean (East Asian Coded Character set, or EACC)

EACC is the code used for storing CJK characters and linking them to related variants for indexing in the OCLC system.

#### Script identifier in records

The client adds the following data to field 066 ꞵc in CJK records to indicate the presence of CJK characters:

- \$1

#### Entering punctuation

- OCLC suggests using the English keyboard to enter punctuation in CJK data fields.
- You may enter CJK punctuation marks using one of the CJK Input Method Editors only if the marks are in the MARC-8 character set (EACC) (see a list of input codes in "Use CJK E-Dictionary—"What is Tsang chieh input code?").
- For searching purposes, CJK punctuation and Latin punctuation are normalized; that is, you can enter punctuation either way and find the same records.

### Romanized data

See the *ALA-LC Romanization Tables* for Chinese, Japanese, and Korean on the Library of Congress Web site:

- Chinese - <http://www.loc.gov/catdir/cpso/romanization/chinese.pdf>
- Japanese - <http://www.loc.gov/catdir/cpso/romanization/japanese.pdf>
- Korean - <http://www.loc.gov/catdir/cpso/romanization/korean.pdf>

### CJK E-Dictionary

See separate topic "Use the CJK E-Dictionary." Open this electronic dictionary from the Tools menu. The CJK dictionary:

- Lets you search or browse to retrieve information about a single character, a group of related characters, homophones matching a phonetic input code, or a large set of characters in sequence by East Asian Character Code (EACC) or by Unicode value.
- Shows comprehensive types of representation for each CJK character supported in Connexion client. Open an entry in the E-Dictionary to copy a value and paste it into a record, workform, constant data, or text string.

### Convert invalid CJK characters to equivalent MARC-8 characters

When you verify CJK characters as MARC-8-compliant (**Edit > MARC-8 Characters > Verify**), and the client identifies invalid character(s), you can automatically convert the character(s) in the record to MARC-8-equivalent CJK characters:

Action
Click <b>Edit &gt; MARC-8 Characters &gt; Convert to MARC-8 CJK</b> , or press <Alt><E><8><J>.
<b>Result:</b> The client converts the characters and changes the color of converted characters to green (by default) or to a color you specify in <b>Tools &gt; Options &gt; Record Display</b> .

**Tip:** If you already know that a record contains invalid CJK characters, you can use the **Edit > MARC-8 Characters > Convert to MARC-8** command without first using the **Edit > MARC-8 Characters > Verify** command.

**Note:** The Library of Congress also has a CJK Compatibility Database on the Cataloging Policy and Support Office (CPSO) home page at [http://www.loc.gov/ils/cjk\\_search/cjk\\_cpso.html](http://www.loc.gov/ils/cjk_search/cjk_cpso.html) to help with MARC-8 compliant or missing characters.

### Use the Chinese, Japanese, or Korean client interface

Change the interface language from English to Chinese (simplified or traditional), Japanese, or Korean. Select the interface language when you:

- Install the Connexion client for the first time or upgrade the client to version 1.30 or higher.
- Open a new user profile you created based on a profile for which you have not already set the interface language.
- Change the interface option at any time in **Tools > Options > International**.

**Note:** To display the Chinese, Japanese, or Korean interface, you must have an input method for the language installed on your workstation (see "Connexion client international cataloging/Input methods for languages that use non-Latin scripts"), or you must have a Chinese, Japanese, or Korean language version of Windows.

### Indexing for CJK script searches

For CJK script searches, the system indexes both single characters and immediately adjacent characters in a field. Use the following search strategies:

- Word search - Enter an index label and a colon (for example **ti:**) followed by a character string with no spaces to find a single word, or followed by more than one character string separated by a space to find multiple words, anywhere in an indexed field.
- Phrase search - Enter an index label and an equal sign (for example, **ti=**) followed by a character string to find exact occurrences, starting with the first character in an indexed field and including each succeeding character. Truncate the character string to find the string followed by any other data without having to enter the entire data string as it appears in a field or subfield.

To truncate, enter an asterisk (\*) at the end of the search string. Enter a minimum of three CJK characters before truncating.

- Phrase browse - Enter the Scan command, an index label, and an equal sign (for example, **sca ti=**) followed by a character string. Phrase browsing scans an index for occurrences of the browse string at the beginning of indexed fields, followed by any other data (automatic truncation).

**Note:** Since all MARC-8 CJK characters are indexed singly, if you browsed for a word, the system would scan for the first character only, and results would not be significant.

### Notes on searching

- If you use qualifiers to limit a search, type them using Latin script.
- Do not use derived searching.
- For searching purposes, CJK punctuation and Latin punctuation are normalized; that is, you can enter punctuation either way and find the same records. If you enter CJK punctuation, the characters must be in the MARC-8 character set (EACC) (see a list of input codes in "Use CJK E-Dictionary—"What is Tsang chieh input code?").
- If you want to retrieve all CJK script records or see sample records, use the "character sets present" WorldCat search index (label **vp:**) with the assigned code **cjk**.

#### To find all CJK script records:

Enter **vp:cjk** in the command line of the Search WorldCat window (**Cataloging > Search > WorldCat**).

**Note:** If a search for all CJK script records alone retrieves too many WorldCat records (limit 1,500 records), you must limit the search and try again.

**Examples:**

**vp:cjk/1991-2**

**vp:cjk and mt:bks**

See general procedures and techniques for searching WorldCat in the Cataloging/Search WorldCat booklet or client Help.

## 7 Use CJK E-Dictionary

### What is the CJK E-Dictionary?

The CJK E-Dictionary (electronic dictionary for Chinese, Japanese, and Korean characters):

- Allows you to search or browse to retrieve information about a CJK character, group of related characters, homophones matching a phonetic input code, or a large set of characters in sequence by EACC or Unicode value.
- Includes all CJK characters represented in the East Asian Character Code (EACC) and supported in the Connexion client.
- For each character, provides the following types of character representation (if applicable):
  - EACC bitmap
  - EACC 3-byte code
  - Unicode font representation
  - Unicode
  - Tsang-chieh input code
  - Wade-Giles input code (if applicable)
  - Pinyin input code (if applicable)
  - McCune-Reischauer input code (if applicable)
  - Modified Hepburn input code (if applicable)

See "Use the search results list" section below for a description of each of these types of character representation

See also, "What is EACC?" and "What is Unicode?" and "What is the Tsang-chieh (TC) input code?" below.

**CJK phonetic input codes.** See the *ALA-LC Romanization Tables* on the Library of Congress Web site for CJK phonetic input codes:

- Chinese (<http://www.loc.gov/catdir/cps/romanization/chinese.pdf>)
- Japanese (<http://www.loc.gov/catdir/cps/romanization/japanese.pdf>)
- Korean (<http://www.loc.gov/catdir/cps/romanization/korean.pdf>)

### Why search or browse the CJK dictionary?

Searching allows you to:

- Find a single character by looking up its EACC code
- Find related characters by searching with a partial EACC code
- Get input codes for a Unicode character that you paste into the dictionary search window
- Find characters that match a Tsang-chieh (TC) code or search using a partial TC code to list characters with one or more common graphic elements
- Find characters that match a phonetic input code

Browsing allows you to:

- View information about all characters encoded in an EACC layer
- View a range of EACC codes within a layer (specify the beginning EACC value to display in the layer)
- Print E-Dictionary content for an EACC layer or part of a layer

**How to search or browse the CJK E-Dictionary**

To ...	Do this ...
Open the E-Dictionary	On the Tools menu click <b>CJK E-Dictionary</b> , or press <Alt><T><E>. (Opens the E-Dictionary Search window.)
<b>Search</b>	
Search by EACC value	<ol style="list-style-type: none"> <li>1. In the E-Dictionary Search window, under <b>Input Method</b>, select <b>EACC</b>.</li> <li>2. In the <b>Search by Input Code</b> box:  <b>Find a character:</b> Type the 3-byte (6-character) EACC code.                      Or  <b>Find related characters:</b> Type two asterisks (**) in place of the first byte (2-character code for the 'Plane' in the EACC code system).</li> <li>3. Click <b>Find</b>.</li> </ol>
Search by Tsang-chieh (TC)	<ol style="list-style-type: none"> <li>1. In the E-Dictionary Search window, under <b>Input Method</b>, select Tsang-chieh (TC).</li> <li>2. In the <b>Search by Input Code</b> box:                      Type the complete TC code for a character. Enter 1 to 5 letters that represent graphic elements of Chinese characters.                      Or                      Type a partial TC code by substituting an asterisk (*) for each letter of the code that you do not know.                      See "What is the Tsang-chieh input code?" below for more information.</li> <li>3. Click <b>Find</b>.</li> </ol>
Search by Unicode character	<ol style="list-style-type: none"> <li>1. Before opening the CJK E-Dictionary, copy a Unicode character from a record-editing window.</li> <li>2. After opening the E-Dictionary, in the E-Dictionary Search window, under <b>Input Method</b>, paste the Unicode character:                      Press &lt;Ctrl&gt;&lt;V&gt; or right-click and on the popup menu click <b>Paste</b>.</li> <li>3. Click <b>Find</b>.</li> </ol>

To ...	Do this ...
Search by phonetic input code	<p>1. In the E-Dictionary Search window, under <b>Input Method</b>, select <b>Wade-Giles</b>, <b>Pinyin</b>, <b>Modified Hepburn</b>, or <b>McCune-Reischauer</b>.</p> <p>2. In the <b>Search by Input Code</b> box, type an input code of the selected type.</p> <ul style="list-style-type: none"> <li>— Wade-Giles and Pinyin (Chinese): Enter 1 to 8 characters. May follow with a tone qualifer and/or TC qualifier.*</li> <li>— Modified Hepburn (Japanese): Enter 1 to 14 characters. May follow with a TC qualifier.*</li> <li>— McCune-Reischauer (Korean): Enter 1 to 8 characters. May follow with a TC qualifier.*</li> </ul> <p>3. Optional. Include diacritics if necessary:</p> <ul style="list-style-type: none"> <li>— Place the cursor before the character associated with the diacritic.</li> <li>— Click <b>ALA Entry</b> and select a diacritic; then click <b>Insert and Close</b>.</li> </ul> <p>4. Click <b>Find</b>.</p> <p>*For tone and Tsang-chieh (TC) qualifiers, enter:</p> <ul style="list-style-type: none"> <li>• <b>Chinese tone qualifiers</b> <ul style="list-style-type: none"> <li>0 (all tones)</li> <li>1 (flat)</li> <li>2 (rising)</li> <li>3 (falling and then rising)</li> <li>4 (falling)</li> <li>5 none</li> </ul> </li> <li>• <b>Japanese and Korean TC qualifiers</b> <ul style="list-style-type: none"> <li>Tone qualifier 0 and then TC qualifier</li> </ul> </li> <li>• <b>TC qualifiers</b> <ul style="list-style-type: none"> <li>Limit retrieval to Chinese-derived characters. Enter the first non-X letter of the TC input code.</li> </ul> </li> </ul>
<b>Browse</b>	
Browse an EACC layer	<p>1. In the E-Dictionary Search window, click <b>Browse</b>.</p> <p>2. In the Browse by EACC Layer window, under <b>Index</b>, select a layer.</p> <p>3. <b>Optional</b>. In the <b>Goto EACC Value</b> box, type a 3-byte (6-character) EACC value for the character with which to start retrieving E-Dictionary entries.</p> <p>4. Click <b>OK</b>.</p> <p><b>Caution:</b> Browsing layer 1 can take up to several minutes; it contains entries for almost 10,000 characters. Layers 3, 8, and 14 are also extensive. See "EACC layers: details" below for the number of characters in each layer.</p>
<b>Search and browse results</b>	
View or print list of matching entries	See "Open and use a dictionary entry" below.
Open a list entry to copy and paste values into records, workforms, constant data, or text strings	See "Open and use a dictionary entry" below.

**EACC layers: details** For basics, see "What is EACC?" below.

- EACC divides the codespace into 16 layers of 6 planes each.
- Layers 10-12 and 15-16 are not currently allocated to characters.
- Within layers, not all planes are in use.
- Layers 1-12 are structured to show relationships between traditional Chinese characters and characters derived from traditional forms (layers 10-12 not in use).
- Layer 1 has traditional characters.
- Layer 2 has simplified forms of characters in layer 1.
- Layers 3 to 9 have other variants of characters in layer 1.

Linked characters and indexing:

- What is an IME (Input Method Editor) and how do I use it?  
(< [http://www.microsoft.com/globaldev/handson/user/IME\\_Paper.mspix](http://www.microsoft.com/globaldev/handson/user/IME_Paper.mspix) >)
- Input Language: Keyboards and IMEs  
(< [http://www.microsoft.com/globaldev/getWR/steps/WRG\\_kybrd.mspix](http://www.microsoft.com/globaldev/getWR/steps/WRG_kybrd.mspix) >)
- The EACC code for a variant form has the same values in the second and third bytes as the code for its corresponding traditional character, but a different value for the first byte.
- Characters with shared second and third byte values are linked.
- In WorldCat, this EACC-code linking is used to index all variants of a character as equivalents.
- A search term in Connexion that begins with a character related by EACC coding to one or more variant forms retrieves records in which the indexed field begins with any of the related forms of that character.

The following table shows the number of characters, EACC planes, and a description for each EACC layer:

Layer	Number of characters	EACC planes	Description
1	9,900	21, 22, 23	Traditional forms of Chinese characters
2	2,100	27, 28, 29	Simplified forms of Chinese traditional characters; linked to characters in layer 1
3	600	2D, 2E, 2F	Variant forms of Chinese traditional characters; linked to characters in layer 1
4	230	33, 34, 35	Other variant forms of Chinese traditional characters; linked to characters in layer 1
5	90	39, 3A, 3B	
6	30	3F	
7	55	45, 46, 47	

Layer	Number of characters	EACC planes	Description
8	480	4B, 4C, 4D	Variant forms of Chinese traditional characters; linked to characters in layer 1; not included in Chinese Character Code for Information Interchange (CCCII)
9	18	51, 52	
10 - 12	Not used		
13	298	69	Japanese characters (hiragana, katakana, and Japanese-invented kokuji) with no Chinese counterparts
14	2,073 (6F) 45 (70)	6F 70	<b>6F</b> - Korean characters (hangul) with no Chinese counterparts  <b>70</b> - Chinese characters created in the People's Republic of China that have no related traditional forms
15 - 16	Not used		




**Use the search results list**

The client displays search and browse results in a list of entries with the following type of representation for each entry (character), if applicable (columns listed in default order).

Character representation	Description
EACC	3-byte East Asian Character Code (EACC) value for the character
Unicode Font (UFont) (graphic)	Character as represented in the default client font  <b>Default:</b> Arial Unicode MS or Windows default  <b>Note:</b> If you change the default client font, the new font must be Unicode-compliant.
Unicode	2-byte code for the character assigned in the Unicode standard
Tsang-Chieh (TC)	One to five letters that identify the character in the Tsang-chieh graphical coding scheme
Wade-Giles (WG)	Wade-Giles romanization of Chinese pronunciation(s) for the character, including any tone qualifier
Pinyin (PY)	Pinyin romanization of Chinese pronunciation(s) for the character; includes any tone qualifier
McCune-Reischauer (MC)	McCune-Reischauer romanization of Korean pronunciation(s) of the character
Modified Hepburn (HP)	Modified Hepburn romanization of Japanese pronunciation(s) of the character

To copy and paste a value or graphic from a dictionary entry, you must open the entry. You cannot copy from the list.

To temporarily customize the list for display and print:

Action
<p><b>Sort order:</b> To re-sort by another column, click the column heading. The list re-sorts by values in the column, in ascending order (alphabetic or numeric) with entries that have no value appearing first. An arrow identifies the sorting column;</p> <p>for example: </p> <p><b>Sort direction:</b> To reverse sort order, click the column heading for the column that sorts the list. The arrow reverses order to indicate descending order:</p> <p></p> <p><b>Column size:</b> To change column size, point to the right border of the column heading you want to re-size. When the pointer becomes , click and hold to drag the border.</p> <p><b>Column order:</b> To change the order of a column, click, hold, and drag the column heading to a new position.</p> <p><b>Note:</b> The client does not retain changes you make to the list once you close the list.</p>

To print the list or print specified pages of the list:

Action
<p>To print an entire results list, click <b>Print</b>. In the Print window, click <b>OK</b>.</p> <p>To print a specified number of pages in the list: When the Print window opens, click the <b>Pages from...to</b> button and enter the number of pages to print. Then click <b>OK</b>.</p> <p><b>Caution:</b> If you want to print results for an entire layer of EACC values, be aware that some layers contain a large number of entries:</p> <ul style="list-style-type: none"><li>• Layer 1 contains entries for almost 10,000 characters. Printing all of the entries requires over 200 sheets of paper.</li><li>• Layers 2, 3, 8, and 14 are also extensive.</li><li>• See "EACC layers: Details" below for the number of characters in each EACC layer.</li></ul>

**Open and use a dictionary entry**

To open a dictionary entry from the E-Dictionary Search Results window:

Action	
	Double-click an entry, or select an entry and press <Enter>.
	The E-Dictionary Entry window opens showing the same information that was in the results list, but in boxes for copying and pasting.

To copy and paste information from a dictionary entry into records, workforms, constant data, or text strings:

	Action
1	In the E-Dictionary Entry window, select a value or graphic.
2	To copy, press <Ctrl><C>, or right-click the selected text or graphic, and on the popup shortcut menu, click <b>Copy</b> .
3	Close the E-Dictionary.
4	To paste, place the cursor where you want paste the text or graphic in a record or text string and press <Ctrl><V>, or right-click where you want to paste, and on the shortcut menu, click <b>Paste</b> .

**What is EACC?**

**Definition.** EACC (East Asian Character Code) is a set of unique 3-byte hexadecimal codes used to represent and store in machine-readable form all Chinese, Japanese, and Korean characters used in the USMARC format.

See East Asian Code Tables in *MARC 21 Specifications for Record Structure, Character Sets, and Exchange Media*. available on the Library of Congress Web site at: < <http://www.loc.gov/mark/specifications/spechome.html> >.

**Standards.** EACC provides codes for characters from the following national coding standards:

- *Chinese Character Code for Information Interchange (CCCII)*, 1982
- *Code of Chinese Graphic Character Set for Information Interchange Primary Set: The People's Republic of China National Standard (GB 2312-80)*
- *Code of the Japanese Graphic Character Set for Information Interchange: Japanese Industrial Standard (JIS C 6226)*, 1983
- *Korean Information Processing System (KIPS)*

**Characteristics of EACC coding structure**

- Based on concept of a 3-dimensional codespace.
- 94 planes, 94 sections in each plane, 94 positions in each section.
- For any EACC code, the first byte designates the plane, the second byte identifies the section within that plane, and the third byte gives the character's position within the section.

- All bytes in hexadecimal; planes, sections, and positions are each numbered from 21 to 7E.
- **Example:** EACC value 214C3C encodes a character on plane 21, in section 4C, at position 3C.

### What is Unicode?

**Definition.** Unicode is the universal character encoding scheme for written characters and text. It defines a consistent way of encoding multi-script text that enables the exchange of text data internationally.

Unicode provides for three encoding forms: a 32-bit form (UTF-32), a 16-bit form (UTF-16), and an 8-bit form (UTF-8, designed for use with ASCII-based systems).

CJK E-Dictionary Unicode values for characters are given in UTF-16 form.

### Unicode Standard, Version 4.0

- Contains 96,382 characters from scripts used for languages worldwide.
- Identical to International Standard ISO/IEC 10646:2003, *Information Technology Universal Multiple-Octet Coded Character Set (UCS) Architecture and Basic Multilingual Plane, Supplementary Planes*, known as the Universal Character Set (UCS)
- In print: *The Unicode Standard, Version 4.0*. The Unicode Consortium. Addison-Wesley Developers Press, 1996.  
Or  
On the Web: < <http://www.unicode.org/versions/Unicode4.0.0/> >.

### CJK characters in Unicode

- CJK unified ideographs contain a set of unified Han ideographic characters used in the written Chinese, Japanese, and Korean languages. The term Han, derived from the Chinese Han Dynasty, refers generally to Chinese traditional culture. The Han ideographic characters make up a coherent script, which was traditionally written vertically, with the vertical lines ordered from right to left. In modern usage, especially in technical works and in computer-rendered text, the Han script is written horizontally from left to right and is freely mixed with Latin or other scripts.
- The unified Han subset contains 70,207 ideographic characters defined by national and industry standards of China, Japan, Korea, Taiwan, Vietnam, and Singapore.
- When used in writing Japanese or Korean, the Han characters are interspersed with other scripts unique to those languages (Hiragana and Katakana for Japanese; Hangul syllables for Korean).
- Unicode includes Hiragana, Katakana, and Hangul script characters.

### EACC character set not fully represented

- CJK characters represented in Unicode are drawn from coding standards developed in China, Japan, Korea, and Taiwan. Characters duplicated in multiple standards are unified and given a single Unicode value. Character forms

assigned separate codes in any primary standard are retained as separate characters in Unicode.

- EACC is among the standards used as sources for Unicode, but is not considered a primary source. Thus, some character forms with unique EACC codes do not have unique Unicode values.
- Unicode lacks codes for approximately 250 variant character forms that have EACC values. Unicode omits these characters as insignificant variants and merges each variant with a similar Unicode form.

The Connexion client maps merged/Unicode-omitted characters to the established Unicode value for display via the Unicode font. Internally, however, the client stores the variants' EACC codes.

### **What is Tsang-chieh? Definition**

- 1 to 5 Latin script letters assigned to represent significant graphic elements of a Chinese character.
- Provides a unique code for each character.
- Can be used as a qualifier with a phonetic input code.

### **Basics**

To use the TC input code, you must know graphic elements of characters, Latin script letters used to represent the graphic elements (see chart below), and rules for selecting significant graphic elements to code (see section below).



Code for a simple character or a compound character segmented by head (outermost, uppermost, or leftmost TC graphic element) and body (remaining graphic elements):

Simple character	Compound character	
	Head	Body
1-4 significant graphic elements: Type a code letter for each	1-2 elements: Type a code letter for each.	1-3 elements: Type a code letter for each
5 or more elements: Type a code letter for each of the first 3 and the last	3 or more elements: Type a code letter for the first and last.	4 or more (simple body): Type a code letter for the first 2 and the last  4 or more (compound body with subhead and subbody):  Subhead 1 - Type a code letter for the first and last elements of the subhead and a letter for the first and last elements of the subbody  Subhead 2 or more - Type a code letter for the first and last elements of the subhead and a letter for the last element of the subbody

- Follow these guidelines to code:
  - Simplicity - Minimize the number of graphic elements. Choose the letter that represents the most inclusive graphic element.
  - Completeness - Within a part or unit, enter the letter for the most inclusive graphic element first.
  - Encircled elements - If a character has more than 5 elements, exclude elements totally encircled by a surrounding unit).
  - Stroke unity - Do not code a letter for an element that retraces part of another element. But do code for each stroke as a unit.
  - Exceptions:**
    - Frequently used characters - Use simplified input codes (including letters for the first and last elements) for certain common characters.
    - Primary input elements - Always enter the following elements first, even if they are not first in traditional calligraphic sequence: 木火大
    - Irregular elements - Use **X** to code for elements not represented by any of the 24 letters defined in standard TC coding.

**Note:** When you search the E-Dictionary by phonetic input code, you can apply the initial Latin script letter assigned to a graphic element in the TC input code (see table above) as a TC qualifier.

**Input codes for CJK punctuation, special characters, and roman punctuation**

To enter the characters and punctuation listed in the following table using one of the CJK Input Method Editors, type the corresponding codes.

Character or punctuation	Description	Input code
•	Center dot	vc
々	Character repetition (Kanji Kurikaeshi kigo)	vk
》	Close angle bracket	v>
】	Close bracket	v]
》》	Close double angle bracket	v'
、	Comma	v,
】	Roman close bracket	]
⋮	Roman colon	:
、	Roman comma	,
—	Roman hyphen	-
[	Roman open bracket	[
⋮	Roman period	.
?	Roman question mark	?
;	Roman semicolon	;
/	Roman slash	/
ー	Long vowel (Choon kigo)	vl ("ell")
《	Open angle bracket	v<
【	Open bracket	v[

Character or punctuation	Description	Input code
◀	Open double angle bracket	v"
。	Period	v.
全	Phrase repetition (ditto mark)	vd
0	Zero	vz

**Note on searching:** For searching purposes, CJK punctuation and Latin punctuation are normalized; that is, you can enter punctuation either way and find the same records.

## 8 Catalog using Cyrillic scripts

### About using Cyrillic scripts

**Cataloging.** Use Cyrillic script data for cataloging items in languages that use the Cyrillic script (for example, Russian, Bulgarian, Serbian, Ukrainian). Use Cyrillic script data the same way you use other non-Latin script data in the client.

See "Use non-Latin scripts for cataloging" for details on using Cyrillic scripts and Cyrillic records. See client Help or other Connexion Client System Guides on the OCLC Web site for client procedures describing how to:

- Search WorldCat
- Create records
- Edit records
- Use constant data
- Save records
- Report errors in records
- Export records
- Import records

**Authorities.** Cyrillic scripts can also be used to add variant name headings to authority records (see section above for more information).

### Tools for using non-Latin scripts

The client provides the following general tools to help you catalog using non-Latin scripts:

- MARC-8 character verification (**Edit > MARC-8 Characters > Verify**) - verify characters separately from record validation.
- Link/unlink fields (**Edit > Linking Fields > Link [or Unlink]**) - visually link non-Latin script data fields with equivalent romanized data fields (bibliographic records only).
- Export options for data fields (**Tools > Options > International**) - determine (for bibliographic records only):
  - Whether to export both Latin-script-equivalent (romanized) data and non-Latin script data or only one or the other
  - Position of data if both
  - Sort order
- Export and import data in UTF-8 Unicode or MARC-8 format. The Unicode option allows you to work with non-MARC-8 characters in the client for your local records. Settings for export are in **Tools > Options > Export**, click **Record Characteristics**, and settings for import are in **File > Import Records**, click **Record Characteristics**.
- See "Use non-Latin scripts for cataloging" for specific procedures for using these tools.

### Cyrillic script entry and character set

#### Script entry methods

- If your system default language is not the one you want to use for cataloging Cyrillic materials, you can install the languages you need in Windows. When you install a language, Windows provides an input keyboard for entering Cyrillic script. See "Connexion client international cataloging," "Input methods for languages that use non-Latin scripts."
- OCLC provides an alternative Cyrillic script keyboard developed for RLIN21 cataloging software. You can download the Cyrillic keyboard from OCLC's Web site. For instructions, see *Getting Started* at <http://www.oclc.org/support/documentation/connexion/client/gettingstarted/gettingstarted>.

See also, *RLIN21 Keyboards* for graphic illustrations of all keyboards, including Hebrew, at <http://www.oclc.org/support/documentation/connexion/client/gettingstarted/gettingstarted/rlin21keyboards.pdf>.

RLIN21 keyboards include characters specific to each script (covering multiple languages that use Arabic script), whereas Microsoft keyboards include script characters specific to a single language.

#### Character sets supported

The client supports the basic and extended Cyrillic character sets defined in *MARC 21 Specifications for Record Structure, Character Sets, and Exchange Media* on the Library of Congress Web site at:

<http://lcweb2.loc.gov/cocoon/codetables/4E.html> (Basic)

<http://lcweb2.loc.gov/cocoon/codetables/51.html> (Extended):

- 4E(hex) [ASCII graphic: N] = Basic Cyrillic
- 51(hex) [ASCII graphic: Q] = Extended Cyrillic

#### Script identifiers in records

The client adds the following data to field 066 ꞗ in Cyrillic records to indicate the presence of Cyrillic characters:

- **(N)** for basic Cyrillic
- **(Q)** for extended Cyrillic

#### Romanized data

See the *ALA-LC Romanization Table* for Cyrillic on the Library of Congress Web site at < <http://www.loc.gov/catdir/cpsol/romanization/nonslav.pdf>.

### Indexing for Cyrillic script searches

#### Notes on searching:

- Use word or phrase search indexes and browse indexes.
- Word searches find the data string you enter anywhere in the indexed field. Phrase searches find the data string starting with the first character in a field or

subfield and including each character in exact order. Browsing scans an index for the closest match to the character string followed by any other data.

- If you use qualifiers to limit a search, type them in Latin script.
- Do not use derived searching.
- Do not truncate searches (asterisk (\*) at the end of a search term). You can use browsing for automatic truncation (enter only as many characters as needed for a match, without using an asterisk at the end).
- If you want to retrieve all Cyrillic script records or see sample records, use the "character sets present" WorldCat search index (label **vp:**) with the assigned code **cyr**.

### **To find all Cyrillic script records:**

Enter **vp:cyr** in the command line of the Search WorldCat window (**Cataloging > Search > WorldCat**).

**Note:** If a search for all Cyrillic script records alone retrieves too many WorldCat records (limit 1,500 records), you must limit the search and try again.

### **Examples:**

**vp:cyr/1991-2**

**vp:cyr and mt:bks**

See general procedures for searching WorldCat in the Cataloging/Search WorldCat booklet or client Help.

### **Cyrillic character indexing specifics:**

All capital or uppercase/capital forms of Cyrillic letters are indexed the same as the corresponding small or lowercase/small forms of Cyrillic letters.

For example, Cyrillic capital letter yu, Ю , is indexed the same as Cyrillic small letter yu, ю .

Type either a capital or small version of a Cyrillic letter in a search and retrieve results for both.

## 9 Catalog using Devanagari script

### About using the Devanagari script

Use Devanagari script data for cataloging items in languages that use the Devanagari script (for example, Hindi, Marathi, Sanscrit, Nepali, Sherpa). Use Devanagari script data the same way you use other non-Latin script data in the client. See "Use non-Latin scripts for cataloging" for details on using Devanagari script data and Devanagari records. See client Help or other Connexion Client System Guides on the OCLC Web site for general procedures describing how to:

- Search WorldCat
- Create records
- Edit records
- Use constant data
- Save records
- Report errors in records
- Export records
- Import records

### Tools for using non-Latin scripts

The client provides the following general tools to help you catalog using non-Latin scripts:

- Link/unlink fields (**Edit > Linking Fields > Link [or Unlink]**) - visually link non-Latin script data fields with equivalent romanized data fields.
- Export options for data fields (**Tools > Options > International**) - determine:
  - Whether to export both Latin-script-equivalent (romanized) data and non-Latin script data or only one or the other
  - Position of data if both
  - Sort order
- **Caution:** MARC-8 character verification (**Edit > MARC-8 Characters > Verify**) is **not** appropriate for verifying Devanagari characters. There is no MARC-8 character set for Devanagari. Using this command for Devanagari results in marking all Devanagari characters as invalid. The OCLC system validates Devanagari characters when you validate a record.

See "Use non-Latin scripts for cataloging" for more specific procedures for working with these tools.

### UTF-8 Unicode export and import required for Devanagari records

Because Devanagari script is not included in MARC-8 character sets, you must export and import records in Unicode format (settings are in **Tools > Options > Export** and in **File > Import Records/Options** button).

### About Unicode

#### Definition

Unicode is the universal character encoding scheme for written characters and text. It defines a consistent way of encoding multi-script text that enables the exchange of text data internationally.

Unicode provides for three encoding forms: a 32-bit form (UTF-32), a 16-bit form (UTF-16), and an 8-bit form (UTF-8, designed for use with ASCII-based systems).

#### Unicode Standard, Version 4.0

- Contains 96,382 characters from scripts used for languages worldwide.
- Identical to International Standard ISO/IEC 10646:2003, *Information Technology Universal Multiple-Octet Coded Character Set (UCS) Architecture and Basic Multilingual Plane, Supplementary Planes*, known as the Universal Character Set (UCS)
- *The Unicode Standard*, Version 4.0. The Unicode Consortium. Addison-Wesley Developers Press, 2003.

### Devanagari script entry and character set

#### Script entry method

If your system default language is not Devanagari, you can install Devanagari, and Windows provides an input keyboard for entering Devanagari script.. See "Connexion client international cataloging/Input methods for languages that use non-Latin scripts."

#### Character set supported

Devanagari characters are defined in Unicode 4.0 (coded in the range U+0900 to U+097F).

#### Script identifier in records

The client adds the following data to field 066 ꣳc in Devanagari records to indicate the presence of Devanagari characters:

- **Deva**

#### Romanized data

See the *ALA-LC Romanization Table* for Devanagari on the Library of Congress Web site at <http://www.loc.gov/catdir/cpsol/romanization/devanagari.pdf>.

### Indexing for Devanagari script searches

#### Notes on searching:

- Use word or phrase search indexes and browse indexes.
- Word searches find the data string you enter anywhere in the indexed field. Phrase searches find the data string starting with the first character in a field or subfield and including each character in exact order. Browsing scans an index for the closest match to the character string followed by any other data.

- If you use qualifiers to limit a search, type them in Latin script.
- Do not use derived searching.
- You can truncate searches (asterisk (\*) at the end of a search term) or use browsing for automatic truncation (enter only as many characters as needed for a match, without using an asterisk).
- If you want to retrieve all Devanagari script records or see sample records, use the "character sets present" WorldCat search index (label **vp:**) with the assigned code **dev**.

### To find all Devanagari script records:

Enter **vp:dev** in the command line of the Search WorldCat window (**Cataloging > Search > WorldCat**).

**Note:** If a search for all Devanagari script records alone retrieves too many WorldCat records (limit 1,500 records), you must limit the search and try again.

### Examples:

**vp:dev/1991-**

**vp:dev and mt:bks**

See general procedures and search techniques for searching WorldCat in the Cataloging, Search WorldCat booklet or client Help.

### Devanagari character indexing specifics:

- The following Devanagari signs are indexed: Candrabindu, Anusvara, Visarga, Nukta, Avagraha, Virama (Halant), and OM (sacred Hindu syllable).
- The Devanagari signs Anudatta (stress sign), Grave accent, and Acute accent are ignored for indexing.
- Independent vowels, dependent vowels, additional consonants, and generic or Devanagari-specific character additions are all indexed as is.  
**Note:** The character for Devanagari vowel short A is unavailable in the Arial Unicode MS font (Connexion client default font).
- Consonants are indexed as is, if attached with Virama (Hasant). Otherwise, they are indexed with the dependent vowel.
- Both Devanagari and Latin numbers are indexed (either may appear in Devanagari text).

### Notes on sorting search results

- Devanagari syllables with candrabindu or anusvara (nasalization signs) precede terms without those syllables.
- Non-conjunct forms of a consonant precede conjunct forms.
- The default sort order for search results—alphabetical sorting by Latin script—is recommended if romanized (Latin-equivalent) data is included in the record. The sort order option is in **Tools > Options > International**.

## 10 Catalog using Greek script

### About using Greek script

**Cataloging.** Use Greek script data for cataloging items in the Greek language. Use Greek script data the same way you use other non-Latin script data in the client. See "Use non-Latin scripts for cataloging" for details on using Greek script and Greek records. See client Help or other Connexion Client System Guides on the OCLC Web site for general procedures describing how to:

- Search WorldCat
- Create records
- Edit records
- Use constant data
- Save records
- Report errors in records
- Export records
- Import records

**Authorities.** Greek script can also be used to add variant name headings to authority records (see section above for more information).

### Tools for using non-Latin scripts

The client provides the following general tools to help you catalog using non-Latin scripts:

- MARC-8 character verification (**Edit > MARC-8 Characters > Verify**) - verify characters separately from record validation.
- Link/unlink fields (**Edit > Linking Fields > Link [or Unlink]**) - visually link non-Latin script data fields with equivalent romanized data fields (for bibliographic records only).
- Export options for data fields (**Tools > Options > International**) - determine (for bibliographic records only):
  - Whether to export both Latin-script-equivalent (romanized) data and non-Latin script data or only one or the other
  - Position of data if both
  - Sort order
- Export and import data in Unicode or MARC-8 format - Unicode option allows you to work with non-MARC-8 characters that are unsupported in the client and in the WorldCat database for your local records (**Tools > Options > Export** and **File > Import Records/Options** button).

### Greek script entry and character set

#### Script entry method

If your system default language is not the one you want to use for cataloging Greek materials, you can install Greek. Windows provides an input keyboard for entering Greek script. See "Connexion client international cataloging/Input methods for

languages that use non-Latin scripts" for more about input methods for non-Latin scripts, or see relevant topics on the Microsoft Web site.

### Character set supported

The client supports the following Greek character set defined in *MARC 21 Specifications for Record Structure, Character Sets, and Exchange Media* on the Library of Congress Web site at <http://lcweb2.loc.gov/cocoon/codetables/53.html>

- 53(hex) [ASCII graphic:S] = Basic Greek

### Script identifier in records

The client adds the following data to field 066 Ꞇc in Greek records to indicate the presence of Greek characters:

- (S

### Romanized data

See the *ALA-LC Romanization Table* for Greek on the Library of Congress Web site at <http://www.loc.gov/catdir/cpsd/romanization/greek.pdf>.

## Indexing for Greek script searches

### Notes on searching:

- Use word or phrase search indexes and browse indexes.
- Word searches find the data string you enter anywhere in the indexed field. Phrase searches find the data string starting with the first character in a field or subfield and including each character in exact order. Browsing scans an index for the closest match to the character string followed by any other data.
- If you use qualifiers to limit a search, type them in Latin script.
- Do not use derived searching.
- Do not truncate searches (asterisk (\*) at the end of a search term). You can use browsing for automatic truncation (enter only as many characters as needed for a match, without using an asterisk at the end).
- If you want to retrieve all Greek script records or see sample records, use the "character sets present" WorldCat search index (label **vp:**) with the assigned code **gre**.

### To find all Greek script records:

Enter **vp:gre** in the command line of the Search WorldCat window (**Cataloging > Search > WorldCat**).

**Note:** If a search for all Greek script records alone retrieves too many WorldCat records (limit 1,500 records), you must limit the search and try again.

### Examples:

**vp:gre/1991-**

**vp:gre and mt:bks**

See general procedures for searching WorldCat in the Cataloging/Search WorldCat booklet or client Help.

### **Greek character indexing specifics:**

All capital forms of Greek letters are indexed the same as the corresponding small forms of Greek letters.

For example, Greek capital letter gamma, Γ, is indexed the same as Greek small letter gamma, γ.

Type either a capital or small version of a Greek letter in a search and retrieve results for both.

Also:

- Greek beta symbol/small letter beta middle of word, β̂, is indexed the same as Greek small letter beta, β.
- Greek small letter final sigma/small letter sigma end of word, σ̂, is indexed the same as Greek small letter sigma, σ.

## 11 Catalog using Hebrew script

### About using Hebrew script

Use Hebrew script data to catalog items in the Hebrew language. Use Hebrew script data the same way you use other non-Latin script data in the client.

See "Use non-Latin scripts for cataloging" for details on using Hebrew scripts and Hebrew records. See client Help or other Connexion Client System Guides on the OCLC Web site for procedures describing how to:

- Search WorldCat
- Create records
- Edit records
- Use constant data
- Save records
- Take OCLC actions
- Export or import records

**Authorities.** Hebrew script can also be used to add variant name headings to authority records (see section above for more information).

### Tools for using non-Latin scripts

The client provides the following general tools to help you catalog using non-Latin scripts:

#### Specific tools to help with Hebrew script cataloging

- Toggle alignment for Arabic or Hebrew script data right-to-left or left-to-right using **View > Align Right** (default: right-to-left)
- Use Unicode formatting characters to control correct display of bidirectional data in Arabic and Hebrew records.
- See procedures below for using these Hebrew-specific tools.

#### Other tools to help with non-Latin script cataloging in general

- MARC-8 character verification (**Edit > MARC-8 Characters > Verify**) - verify characters separately from record validation.
- Link/unlink fields (**Edit > Linking Fields > Link [or Unlink]**) - visually link non-Latin script data fields with equivalent romanized data fields (for bibliographic records only).
- Export options for data fields (**Tools > Options > International**) - determine (for bibliographic records only):
  - Whether to export both Latin-script-equivalent (romanized) data and non-Latin script data or only one or the other
  - Position of data if both
  - Sort order

- Export and import data using UTF-8 Unicode or MARC-8 character sets. The UTF-8 Unicode option allows you to work with non-MARC-8 characters in the client for your local records (settings for export are in **Tools > Options > Export**, click **Record Characteristics**, and settings for import are in **File > Import Records**, click Record Characteristics).
- See "Use non-Latin scripts for cataloging" for more specific procedures for working with these tools.

### Hebrew script entry and character set

#### Script entry methods

- If your system default language is not Hebrew, you can install the Hebrew language in Windows. When you install Hebrew, Windows provides an input keyboard for entering Hebrew script. See "Connexion client international cataloging," "Input methods for languages that use non-Latin scripts."
- OCLC provides an alternative Hebrew script keyboard developed for RLIN21 cataloging software. You can download the Arabic keyboard from the OCLC Web site.

See also, *RLIN21 Keyboards* for graphic illustrations of all keyboards, including Hebrew, at <http://www.oclc.org/support/documentation/connexion/client/gettingstarted/gettingstarted/rlin21keyboards.pdf>.

RLIN21 keyboards include characters specific to each script (covering multiple languages that use Arabic script), whereas Microsoft keyboards include script characters specific to a single language.

#### Character set supported

The client supports the following Hebrew character set defined in *MARC 21 Specifications for Record Structure, Character Sets, and Exchange Media*, available on the Library of Congress Web site at: <http://lcweb2.loc.gov/cocoon/codetables/32.html>:

- 32(hex) [ASCII graphic: 2] = Basic Hebrew

#### Script identifier in records

The client adds the following data to field 066 ꞑc in Hebrew records to indicate the presence of Hebrew characters:

- (2

#### Romanized data

See the *ALA-LC Romanization Table* for Hebrew on the Library of Congress Web site at: <http://www.loc.gov/catdir/cpsd/romanization/hebrew.pdf>.

### Align Hebrew or Arabic script data for display and print

By default, the client displays (and prints) Hebrew or Arabic script with data aligned to the right. To toggle between displaying these scripts right-to-left or left-to-right:

Action
<p><b>Toggle alignment for all Hebrew or Arabic script data in the current record:</b> Click <b>View &gt; Align Right</b>, or press &lt;Alt&gt;&lt;V&gt;&lt;I&gt;. <b>Default:</b> Data aligns to the right for display and printing.</p> <p><b>Result:</b> The Align Right icon next to the command on the View menu is active (highlighted) if Align Right is selected. The icon is inactive (grayed out) if Align Right is cleared.</p> <p>Or</p> <p><b>Toggle data alignment in the current field:</b> Right-click a field, and on the pop-up menu, click <b>Right-to-Left Reading Order</b>.</p> <p><b>Result:</b> The client changes alignment of the Hebrew or Arabic script data only in the current field.</p>

### Use Unicode formatting characters to control bidirectional data

Enter Unicode formatting characters in Arabic, Persian, and Hebrew records to correctly display left-to-right multiple-digit numbers and punctuation, including brackets, hyphens, internal spaces, etc., within a field of right-to-left script data.

- **Export/import using UTF-8 Unicode character set.** Unicode formatting control characters are retained as is in Arabic, Persian, and Hebrew records exported or imported using the UTF-8 Unicode character set, along with other non-MARC-8 Unicode characters.
- **Export/import using MARC-8 character set.** The Unicode formatting characters are retained in Numeric Character Reference (NCR) format in records exported or imported using the MARC-8 character set, along with other non-MARC-8 characters.

	<b>Action</b>
1	Click to locate the cursor in the position where you want to insert a formatting control number.
2	Right-click in the field, and on the pop-up menu click <b>Insert Unicode Control Character</b> . Or Right-click to open the pop-up menu and then press the keystrokes shown in step 3.
3	<p>Click one of the following characters (or press the keystroke shortcuts, shown in parentheses):</p> <ul style="list-style-type: none"> <li>• <b>LRM Left-to-Right Mark</b> (&lt;Alt&gt;&lt;R&gt;&lt;L&gt;)</li> <li>• <b>RLM Right-to-Left Mark</b> (&lt;Alt&gt;&lt;R&gt;&lt;R&gt;)</li> <li>• <b>ZWJ Zero Width Joiner</b> (&lt;Alt&gt;&lt;R&gt;&lt;J&gt;)</li> <li>• <b>ZWNJ Zero Width Non-Joiner</b> (&lt;Alt&gt;&lt;R&gt;&lt;N&gt;)</li> <li>• <b>LRE Start of Left-to-Right Embedding</b> (&lt;Alt&gt;&lt;R&gt;&lt;S&gt;)</li> <li>• <b>RLE Start of Right-to-Left Embedding</b> (&lt;Alt&gt;&lt;R&gt;&lt;T&gt;)</li> <li>• <b>LRO Start of Left-to-Right Override</b> (&lt;Alt&gt;&lt;R&gt;&lt;A&gt;)</li> <li>• <b>RLO Start of Right-to-Left Override</b> (&lt;Alt&gt;&lt;R&gt;&lt;F&gt;)</li> <li>• <b>PDF Pop Directional Formatting</b> (&lt;Alt&gt;&lt;R&gt;&lt;P&gt;)</li> </ul> <p><b>Tip for one-step entry:</b> Create a text string using <b>Tools &gt; Text Strings</b>; click <b>Add</b> and enter one of the characters listed above using the right-click menu. Then use the Text Strings quick tool on the toolbar to enter the character.</p> <p>Or</p> <p>Assign the text string to a keystroke shortcut. Enter the character by pressing the keystroke.</p>

**Example:** To control the display of the data *742[1981 or 1982]* that you enter in field 260 Ꞇc, and that is preceded and to be followed by Arabic script data:

1. Click to locate the cursor in field 260 Ꞇc.
2. Right-click in the field, and in the pop-up menu click **Insert Unicode Control Character**. Then click **LRE Start of Left-to-Right Embedding**.
3. Enter the data string, **742[1981 or 1982]**, immediately following the character.
4. Without moving the cursor, right-click in the field again. In the pop-up menu click **Insert Unicode Control Character**. Then click **PDF Pop Directional Formatting**.

**More information:**

- For details, see the *Bidirectional Algorithm* report on the Unicode Web site at: <http://unicode.org/reports/tr9/>
- See more about selecting a character set for exporting and importing bibliographic records in the client in Cataloging, Export or Import Bibliographic Records.

**Indexing for Hebrew script searches**

**Notes on searching:**

- Use word or phrase search indexes and browse indexes.
- Word searches find the data string you enter anywhere in the indexed field. Phrase searches find the data string starting with the first character in a field or subfield and including each character in exact order. Browsing scans an index for the closest match to the character string followed by any other data.
- If you use qualifiers to limit a search, type them in Latin script.
- Do not use derived searching.
- Do not truncate searches (asterisk (\*) at the end of a search term). You can use browsing for automatic truncation (enter only as many characters as needed for a match without using an asterisk).
- If you want to retrieve all Hebrew script records or see sample records, use the "character sets present" WorldCat search index (label **vp:**) with the assigned code **hbr**.

**To find all Hebrew script records:**

Enter **vp:hbr** in the command line of the Search WorldCat window (**Cataloging > Search > WorldCat**).

**Note:** If a search for all Hebrew script records alone retrieves too many WorldCat records (limit 1,500 records), you must limit the search and try again.

**Examples:**

**vp:hbr/1991-  
vp:hbr and mt:bks**

See general procedures for searching WorldCat in the Cataloging/Search WorldCat booklet or client Help.

**Hebrew character indexing specifics:**

The following table shows Hebrew characters grouped together and indexed the same as if they are the same character (the characters are "normalized").

Type any character of a group of normalized characters in a search and retrieve results for all characters in the group.

Images and names of characters indexed the same are in columns 3 and 4, opposite the character with which they are indexed.

Character	Character name	Other characters indexed the same	
		Character	Character name
ו	Vav	וו	Ligature Yiddish double Vav/ Tsvey Vovn
		וּ	Ligature Yiddish Vav Yod/Vov Yud

Character	Character name	Other characters indexed the same	
		Character	Character name
י	Yod	ײ	Ligature Yiddish double Yod/ Tsvey Yudn
כ	Kaf	ך	Final Kaf
מ	Mem	ם	Final Mem
נ	Nun	ן	Final Nun
פ	Pe	ף	Final Pe
צ	Tsadi	ץ	Final Tsadi

## 12 Catalog using Tamil script

### About using Tamil script

Use Tamil script data to catalog items in the Tamil language. Use Tamil script data the same way you use other non-Latin script data in the client. See "Use non-Latin scripts for cataloging" for details on using Tamil scripts and Tamil records. See client Help or other Connexion Client System Guides on the OCLC Web site for procedures describing how to:

- Search WorldCat
- Create records
- Edit records
- Use constant data
- Save records
- Report errors in records
- Export records
- Import records

### Tools for using non-Latin scripts

The client provides the following general tools to help you catalog using non-Latin scripts:

- Link/unlink (**Edit > Linking Fields**) - visually link non-Latin script data fields with equivalent romanized data fields.
- Export options for data fields (**Tools > Options > International**) - determine:
  - Whether to export both Latin-script-equivalent (romanized) data and non-Latin script data or only one or the other
  - Position of data if you export both Latin and non-Latin script data
  - Sort order
- **Caution: MARC-8 character verification (Edit > MARC-8 Characters > Verify)** is **not** appropriate for verifying Tamil characters. There is no MARC-8 character set for Tamil. Using this command for Tamil results in marking all Tamil characters as invalid. The OCLC system validates Tamil characters when you validate a record.

### Unicode export and import required for Tamil records

Because Tamil script is not included in MARC-8 character sets, you must export and import records in Unicode format (settings are in **Tools > Options > Export** and **File > Import Records/Options** button).

### About Unicode

**Definition.** Unicode is the universal character encoding scheme for written characters and text. It defines a consistent way of encoding multi-script text that enables the exchange of text data internationally.

Unicode provides for three encoding forms: a 32-bit form (UTF-32), a 16-bit form (UTF-16), and an 8-bit form (UTF-8, designed for use with ASCII-based systems).

### Unicode Standard, Version 4.0

- Contains 96,382 characters from scripts used for languages worldwide.
- Identical to International Standard ISO/IEC 10646:2003, *Information Technology Universal Multiple-Octet Coded Character Set (UCS) Architecture and Basic Multilingual Plane, Supplementary Planes*, known as the Universal Character Set (UCS)
- In print: *The Unicode Standard, Version 4.0*. The Unicode Consortium. Addison-Wesley Developers Press, 1996.  
Or  
On the Web: < <http://www.unicode.org/versions/Unicode4.0.0/> >.

### Tamil script entry and character set

#### Script entry method

If your system default language is not Tamil, you can install the Tamil language in Windows. When you install Tamil, Windows provides an input keyboard for entering Tamil script. See "Connexion client international cataloging/Input methods for languages that use non-Latin scripts" for more about input methods for non-Latin scripts.

#### Character set supported

Tamil characters are defined in Unicode 4.0 (coded in the range U+0B80 to U+0BFF).

Unicode 4.0 coding for Tamil characters does not include most glyphs that are formed when consonants alone (implicit "a" suppressed) are combined with independent vowels.

**Caution:** Microsoft Arial Unicode MS font, which is recommended for general use in the client, does not support the Tamil digits, numerics, and symbols that are coded in Unicode 4.0.

#### Script identifier in records

The client adds the following data to field 066 ꜥc in Tamil records to indicate the presence of Tamil characters:

- **TamI**

To find Tamil script records in WorldCat, enter the search **vp:tam**. (Enter as a command line search in **Cataloging > Search > WorldCat**.)

#### Romanized data

See the *ALA-LC Romanization Table* for Tamil on the Library of Congress Web site at <http://www.loc.gov/catdir/cpsd/romanization/tamil.pdf>.

### Indexing for Tamil searches

#### Notes on searching:

- Use word or phrase search indexes and browse indexes.
- Word searches find the data string you enter anywhere in the indexed field. Phrase searches find the data string starting with the first character in a field or subfield and including each character in exact order. Browsing scans an index for the closest match to the character string followed by any other data.
- If you use qualifiers to limit a search, type them in Latin script.
- Do not use derived searching.
- Do not truncate searches (asterisk (\*) at the end of a search term). You can use browsing for automatic truncation (enter only as many characters as needed for a match without using an asterisk).
- If you want to retrieve all Tamil script records or see sample records, use the "character sets present" WorldCat search index (label **vp:**) with the assigned code **tam**.

#### To find all Tamil script records:

Enter **vp:tam** in the command line of the Search WorldCat window (**Cataloging > Search > WorldCat**).

**Note:** If a search for all Tamil script records alone retrieves too many WorldCat records (limit 1,500 records), you must limit the search and try again.

#### Examples:

**vp:tam/1991-**  
**vp:tam and mt:bks**

See general procedures for searching WorldCat in the Cataloging/Search WorldCat booklet or client Help.

#### Tamil character indexing specifics:


Tamil independent vowels are characters that stand on their own. Each has a unique Unicode code value.


Tamil consonants contain an implicit "a" vowel sound. A modifier called Virama (or Pulli) added to the consonant glyph represents the consonant alone with no vowel. If the consonant alone is combined with an independent vowel (not implicit "a"), the vowel becomes dependent and the visual form changes.



The OCLC system indexes each Tamil consonant separately in all three of the following forms:

- Consonant with the implicit "a"
- Consonant alone (indicated by modifier Virama or Pulli)
- Consonant combined with any independent vowel (indicated by the addition of a vowel ligature)

**Example:**

- க = [ka] = Tamil consonant KA with implicit "a."
- க் = [k] = consonant KA alone (with Virama or Pulli that looks like a diacritical mark).
- கி = [ki] = consonant KA combined with independent vowel ி,  [i].

The glyph for KA has the added ligature for ி:  (where the dotted circle represents the consonant to which the ligature is attached). Unicode 4.0 does not provide a unique code for the transformed glyph. For example, it does not

code  . It codes only the dependent vowel ligature (for example, .

- Based on Unicode Tamil character-rendering rules, the OCLC system stores the Unicode values in a sequence that represents a new glyph used in Tamil text writing and indexes accordingly. For the example here, the system indexes the glyphs with the following Unicode values:

— க [ka] as 0B95

— க் [k] as 0B95+0BCD, where Virama (0BDC) is indexed

— கி [ki] as 0B95+0BBF

**Note on sorting WorldCat search results:** Tamil Unicode 4.0 codes are not in collating order. The default sort order for search results, alphabetical sorting by Latin script, is recommended if romanized (Latin-equivalent) data is included in the record. The sort order option is in **Tools > Options > International** (or <Alt><T><O>).

## 13 Catalog using Thai script

### About using Thai script

Use Thai script data to catalog items in the Thai language. Use Thai script data the same way you use other non-Latin script data in the client. See "Use non-Latin scripts for cataloging" for details on using Thai scripts and Thai records. See client Help or other Connexion Client System Guides on the OCLC Web site for procedures describing how to:

- Search WorldCat
- Create records
- Edit records
- Use constant data
- Save records
- Report errors in records
- Export records
- Import records

### Tools for using non-Latin scripts

The client provides the following general tools to help you catalog using non-Latin scripts:

- Link/unlink (**Edit > Linking Fields**) - visually link non-Latin script data fields with equivalent romanized data fields (see "Use non-Latin scripts for cataloging").
- Export options for data fields (**Tools > Options > International**) - determine:
  - Whether to export both Latin-script-equivalent (romanized) data and non-Latin script data or only one or the other
  - Position of data if you export both Latin and non-Latin script data
  - Sort order
- **Caution: MARC-8 character verification (Edit > MARC-8 Characters > Verify)** is **not** appropriate for verifying Thai characters. There is no MARC-8 character set for Thai. Using this command for Thai results in marking all Thai characters as invalid. The OCLC system validates Thai characters when you validate a record.

### Unicode export and import required for Thai records

Because Thai script is not included in MARC-8 character sets, you must export and import records in Unicode format (settings are in **Tools > Options > Export and File > Import Records/Options** button).

### About Unicode

**Definition.** Unicode is the universal character encoding scheme for written characters and text. It defines a consistent way of encoding multi-script text that enables the exchange of text data internationally.

Unicode provides for three encoding forms: a 32-bit form (UTF-32), a 16-bit form (UTF-16), and an 8-bit form (UTF-8, designed for use with ASCII-based systems).

### Unicode Standard, Version 4.0

- Contains 96,382 characters from scripts used for languages worldwide.
- Identical to International Standard ISO/IEC 10646:2003, *Information Technology Universal Multiple-Octet Coded Character Set (UCS) Architecture and Basic Multilingual Plane, Supplementary Planes*, known as the Universal Character Set (UCS)
- In print: *The Unicode Standard, Version 4.0*. The Unicode Consortium. Addison-Wesley Developers Press, 1996.  
Or  
On the Web: < <http://www.unicode.org/versions/Unicode4.0.0/> >.

### Thai script entry and character set

#### Script entry method

If your system default language is not Thai, you can install the Thai language. When you install Thai, Windows provides an input keyboard for entering Thai script. See more about input methods for non-Latin scripts in "Connexion client international cataloging/Input methods for languages that use non-Latin scripts."

#### Character set supported

Thai characters are defined in Unicode 4.0 (coded in the range U+0E00 to U+0E7F).

#### Script identifier in records

The client adds the following data to field 066 ꜥc in Thai records to indicate the presence of Thai characters:

- **Thai**

#### Romanized data

See the *ALA-LC Romanization Table* for Thai on the Library of Congress Web site at <http://www.loc.gov/catdir/cpsd/romanization/thai.pdf>.

### Indexing for Thai script searches

For Thai script searches, the system treats the entire data string you enter as both a word and a phrase, since Thai text has no spaces between words. Search for Thai terms using word or phrase search indexes and word or phrase browse indexes.

You must use truncation to search a phrase index and retrieve the exact data string without having to enter the entire data string as it appears in a field or subfield. Truncation retrieves the data string followed by any other data.

- To truncate, enter an asterisk (\*) at the end of the character string. You must enter a minimum of three Thai characters before truncating.

Or use browsing for automatic truncation (enter only as many characters as needed for a match, without using an asterisk at the end).

**Notes on searching:**

- If you use qualifiers to limit a search, type them using Latin script.
- Do not use derived searching.
- For Thai, the system finds the search term as a word, a data string occurring anywhere in a field, or as a phrase, a complete data string starting with the first character in a field or subfield and including each character in the string in exact order, depending on whether you use word indexes (for example, **ti**:[search string]) or phrase indexes (for example, **ti**=[search string]).
- Browsing scans an index for the closest match anywhere in an indexed field for word indexes (for example **sca ti**:[search string]) or at the beginning of any indexed field, followed by any data, for phrase indexes (for example, **sca ti**=[search string]).
- If you want to retrieve all Thai script records or see sample records, use the "character sets present" WorldCat search index (label **vp**;) with the assigned code **tha**.

**To find all Thai script records:**

Enter **vp:tha** in the command line of the Search WorldCat window (**Cataloging > Search > WorldCat**).

**Note:** If a search for all Thai script records alone retrieves too many WorldCat records (limit 1,500 records), you must limit the search and try again.

**Examples:**

**vp:tha/1991-**  
**vp:tha and mt:bks**

See general procedures for searching WorldCat in the Cataloging/Search WorldCat booklet or client Help.

**Thai character indexing specifics:**

The following table shows Thai characters indexed more than one way or indexed as a space:

Character	Character name	How character is indexed in OCLC system
๑	Lo Ling	Indexed, <b>except:</b> If it appears between two Paiyannoi characters ( <b>๑</b> ), the entire three-character string is indexed as a space.
๑	Paiyannoi	Indexed as a space.

International Cataloging: Use Non-Latin Scripts

Character	Character name	How character is indexed in OCLC system
๕	Sara A	Indexed, <b>except</b> : Not indexed if preceded by the character Angkhankhu ( ๕ ).
฿	Baht	Indexed as a space.
๑	Maiyamok	Indexed as a space.
๕	Yamakkan	Indexed as a space.
๐	Fongman	Indexed as a space.
๕	Angkhankhu	Indexed as a space.
๐๗๗	Khomut	Indexed as a space.

## 14 Install RLIN21 Arabic, Cyrillic, Hebrew or Latin keyboards

About keyboards for entering script data in the Connexion client

- **RLIN21 Arabic, Cyrillic, and Hebrew keyboards:**
  - RLIN21 keyboards.** RLIN21 Arabic, Cyrillic, and Hebrew keyboards were previously created by RLG for use with RLIN21. The Connexion client supports the keyboards to assist RLIN21 catalogers switch to the client.
  - Windows input keyboards.** Previously, Connexion client catalogers used Microsoft Windows input keyboards to enter these scripts. Now, you can continue using Windows keyboards or install and use RLIN21 keyboards.
  - Difference:** RLIN21 keyboards include characters specific to each script. Microsoft keyboards include characters specific to a language written in a particular script.

**Example:** Using the single RLIN21 Arabic keyboard, generate Arabic script characters for Arabic, Persian, and Urdu instead of switching among separate Microsoft keyboards for each language to enter script data.

**Note:** For Chinese, Japanese, and Korean (CJK), you must use Windows Input Method Editors (IMEs). For other supported scripts (Bengali, Devanagari, Greek, Tamil, and Thai), use Windows keyboards. See also, Cataloging, International, “Input methods for languages that use non-Latin scripts.”

- **RLIN21 Latin keyboard:**
  - RLIN21 catalogers making the transition to Connexion client who want to enter diacritics and special characters the same way they did in RLIN21 can use the RLIN21 Latin keyboard or RLIN21 equivalent keystroke shortcuts (see the URL location of a list of these keys at the end of this section).
  - Existing client function.** To use the existing functionality for entering a diacritic or special character, click **Edit > Enter Diacritics** and select the image or name for a character, or use assigned client keystrokes (see a complete list of client assigned keys in **View > Assigned Keys**).
  - The client supports both methods of entering diacritics and special characters.

Install an RLIN21 keyboard

**Notes:**

- You must have local administrator rights for your workstation to download and install the RLIN21 keyboards.
- After installing, you must make an installed keyboard available on your workstation—that is, select it under the **English (United States)** input locale for your workstation—before you can use the keyboard.

To install a keyboard:

	Action
1	In your Internet browser, go to <a href="http://psw.oclc.org">http://psw.oclc.org</a> , and click <b>Software downloads</b> in the list of links on the left.
2	Type your OCLC authorization and password and click <b>Enter</b> .

	Action
3	On the <b>Software downloads</b> page, click the link for one of the following: <ul style="list-style-type: none"> <li>• <b>RLIN21 Arabic keyboard</b> (downloads <b>ARBr21.exe</b> file)</li> <li>• <b>RLIN21 Cyrillic keyboard</b> (downloads <b>CYRr21.exe</b> file)</li> <li>• <b>RLIN21 Hebrew keyboard</b> (downloads <b>HBRr21.exe</b> file)</li> <li>• <b>RLIN21 Latin keyboard</b> (downloads <b>LATr21.exe</b> file)</li> </ul>
4	Scroll to the end of the OCLC Software License Agreement and click <b>I Agree</b> .
5	In the <b>File Download</b> window, check the <b>*.exe</b> file name to be sure you are downloading the keyboard you want, and then click <b>Save</b> .
6	In the OCLC Software License Agreement window, click <b>I Agree</b> again to confirm your acceptance of the agreement.
7	In the <b>Save as</b> window, expand the <b>Save in</b> list if needed to navigate to the folder where you want to save the file. Select the folder, and then click <b>Save</b> .  The window title changes to <b>Download Complete</b> . A message in the window also confirms that the file is downloaded.  You can now close the browser.
8	In Windows Explorer, navigate to the folder where you saved the <b>*.exe</b> file. Double-click the file or highlight it and press <Enter>.  You receive an <i>Installation Complete</i> message. Click <b>Close</b> .
9	Repeat this procedure to install additional RLIN21 keyboards if needed.
10	Shut down and restart your workstation before you use the keyboard(s).

**Make an installed RLIN21 keyboard available for use**

After installing a keyboard, you must make it available for use in Windows. Instructions differ somewhat, depending on your version of Windows. Follow instructions on the Microsoft Web site for your version. See:

- Enabling International Support in Windows 2000 at:  
<http://www.microsoft.com/globaldev/handson/user/2kintl supp.msp x>
- Enabling International Support in Windows XP at:  
<http://www.microsoft.com/globaldev/handson/user/xpintl supp.msp x>
- For Windows Vista, see information about languages, scripts, keyboards, and input methods from Microsoft's Vista Global Development and Computing Portal Web page at <http://www.microsoft.com/globaldev/vista/vistahome.msp x>.



**Caution:** You must make RLIN21 keyboards available for use under the **English (United States)** input locale when you follow the Microsoft instructions. Do not install under any other language.


For example, in Windows 2000, follow these steps:

1. Go to the Windows **Start** menu and click **Settings > Regional Options**.
2. Click the **Input Locales** tab. (Depending on your iteration of Windows 2000, that is, the service pack installed, you may need to click **Change** to open an input locales settings dialog.)



3. Under **Installed Services**, click **EN English (United States)** (selected as the default language unless you changed this setting), and then click **Add**.
4. In the Add Input Locale dialog, make sure English (United States) is selected in the **Input Locale** list. Then click to expand the **Input Keyboard/IME** list and scroll to locate the RLIN21 keyboard. Click to select it.  
Or  
If you clicked **Change** and the Add Input Language dialog opened, click the **Keyboard Layout/IME** check box and scroll the list below it to find the RLIN21 keyboard you want to make available. Click to select it.
5. Click **OK**. The RLIN21 keyboard you selected now appears under English (United States) in the Installed Services list of the Input Locales tab.
6. Click **OK**.

**Select an RLIN21 keyboard to input script data**

When you install languages and/or input keyboards or IMEs on your workstation, a language indicator (  ) appears in the Windows system tray (bottom right corner of the desktop). If you have multiple languages installed, clicking the language indicator (  ) expands a list of them. If not, the indicator contains only English (United States).

When you install keyboards under a language and the language is selected in the language indicator, a keyboard indicator (  ) also appears in the system tray.

To select an RLIN21 keyboard for entering data:

	Action
1	<p>Make sure that <b>EN</b> is selected in the language indicator (  ) if necessary (that is, if you have multiple languages installed on your workstation).</p> <p><b>Tip:</b> If you have multiple languages installed, use the Windows default keystroke shortcut <b>&lt;Left Alt&gt;&lt;Shift&gt;</b> to toggle through the languages if needed.</p>
2	<p>Click the keyboard indicator (  ) and select the name of the RLIN21 keyboard you want to use.</p>

**Caution:**

- While using the RLIN21 keyboards, you cannot use Connexion client default keystroke shortcuts mapped to characters that include the **Alt**, **Ctrl+Alt**, or **Ctrl+Alt+Shift** keys.
- These key combinations are used with the RLIN21 keyboards and override the Connexion client keystrokes when you are using an RLIN21 keyboard.
- However, Connexion client function keys and keystrokes assigned to characters that include the **Ctrl**, **Alt+Shift** and **Ctrl+Shift** keys remain available.

**Examples:**

The default keystroke for **Action > Delete Record**, mapped to <Ctrl><Alt><D>, conflicts with the RLIN21 keyboards (contains **Ctrl+Alt**) and cannot be used. You must map **Delete Record** to another keyboard shortcut.

However, you can continue to use the default keystroke for the delimiter character (<Ctrl><D>) and the default keystroke for **Action > Export** (<F5>), since they do not conflict with RLIN21 keyboards.

**Uninstall an RLIN21 keyboard**

**Notes:**

- You must have local administrator rights to uninstall the RLIN21 keyboards.
- You must first remove the keyboard from regional options before uninstalling (next procedure). If you do not, when you uninstall, you receive an error message, Fatal error during installation.

First, remove the keyboard from regional options:

	Action
1	<p><b>Windows 2000:</b></p> <ol style="list-style-type: none"> <li>1. Click the Windows <b>Start</b> menu, and then click <b>Settings &gt; Control Panel &gt; Keyboard</b>.</li> <li>2. Click the <b>Input Locales</b> tab.</li> </ol> <p><b>Windows XP:</b></p> <ol style="list-style-type: none"> <li>1. Click the Windows <b>Start</b> menu, and then click <b>Control Panel &gt; Regional and Language Options</b>.</li> <li>2. Click the <b>Languages</b> tab.</li> <li>3. Click <b>Details</b>.</li> </ol>
2	In the list of language services available, under <b>English (United States)</b> , select the name of the RLIN21 keyboard you want to remove.
3	<p><b>Windows 2000:</b></p> <ul style="list-style-type: none"> <li>• Click <b>Remove</b> and then click <b>OK</b>.</li> </ul> <p><b>Windows XP:</b></p> <ul style="list-style-type: none"> <li>• Click <b>Apply</b> and then click <b>OK</b>.</li> </ul>

Then, uninstall the keyboard program:

	Action
1	Click the Windows <b>Start</b> menu, and then click <b>Settings &gt; Control Panel &gt; Add/Remove Programs</b> .
2	In the alphabetical list of installed programs, scroll to the name of the RLIN21 keyboard that you want to uninstall and select it.
3	Click <b>Remove</b> . Click <b>Yes</b> to confirm.

**Guides - RLIN21  
keyboard layouts and  
RLIN21 Latin  
character keystroke  
shortcuts**

- *RLIN21™ Keyboards* (graphic layouts) at <http://www.oclc.org/support/documentation/connexion/client/gettingstarted/gettingstarted/rlin21keyboards.pdf>
- *RLIN21 Latin Character Key Shortcuts* (directions and a list of keystroke shortcuts for the RLIN21 Latin keyboard) at <http://www.oclc.org/support/documentation/connexion/client/gettingstarted/gettingstarted/rlin21latincharacterkeystrokes.pdf>