SIL Encoding Converters 3.1

Overview

This package provides tools through which you can change the encoding, font, and/or script of text in Microsoft Word documents, XML documents, and SFM text and lexicon documents. It also installs a system-wide repository to manage your encoding converters and transliterators (TECkit, CC, ICU, Perl, or Python based, as well as support for adding custom transduction engines).

For developers, it provides a simple COM interface to select and use a converter from the repository. It is easy to use from VBA, C++, C#, Perl, Python or any .NET/COM enabled language. This package is fully integrated with SIL FieldWorks, Adapt It, and the forthcoming SpeechAnalyzer software, providing the same system-wide registry of installed and available encoding converters for all of these user programs. Additionally the package includes some extra utilities such as a clipboard converter for manipulating text between cut and paste operations.

The following picture illustrates the suite of tools, utilities, and applications that are available and how they interact:

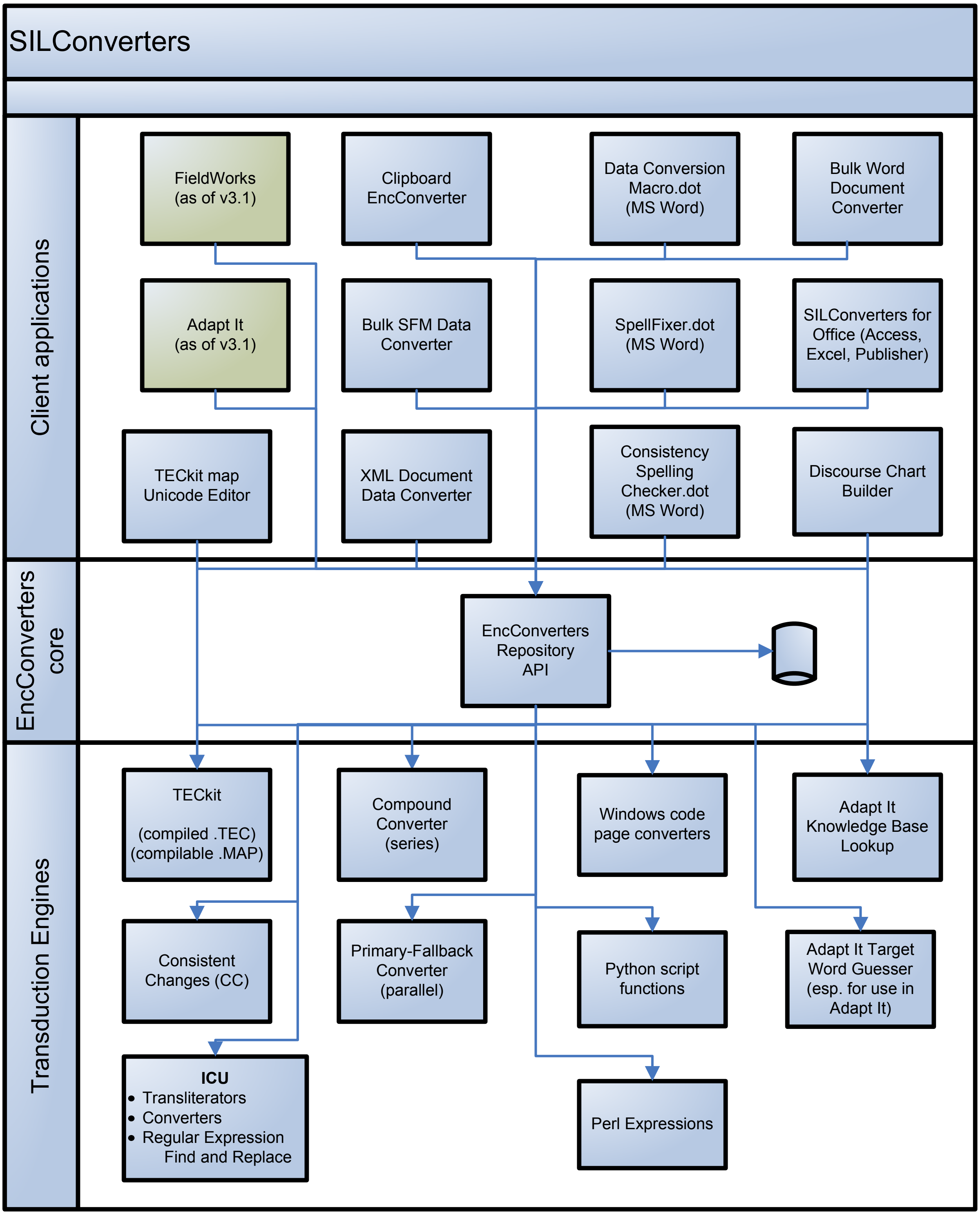
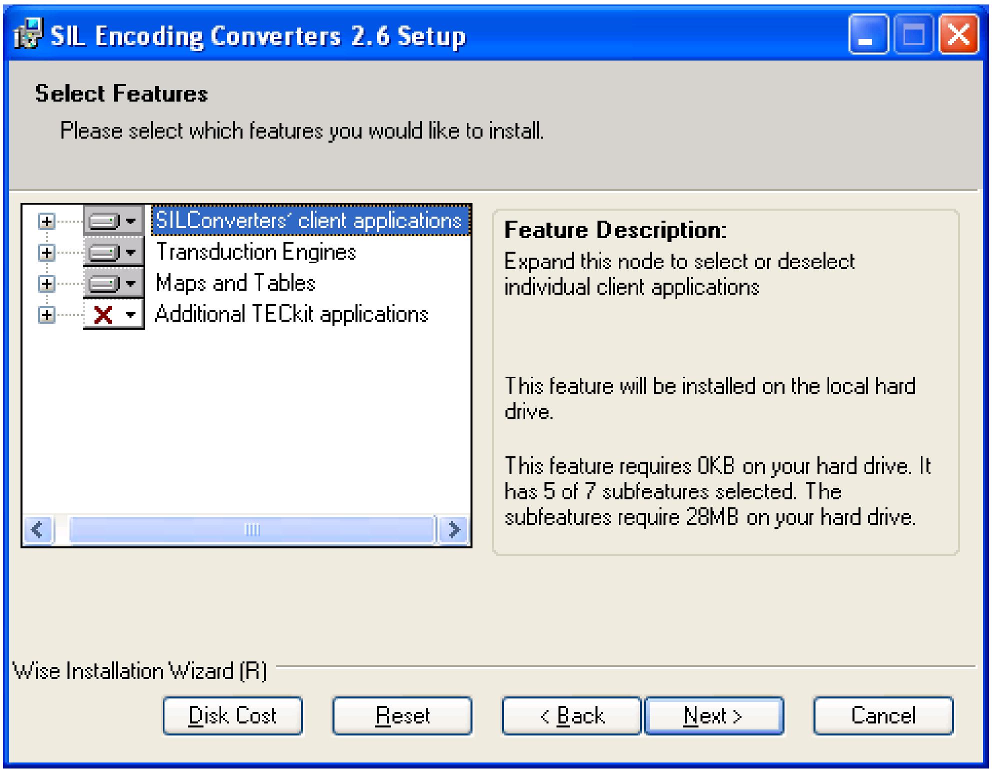

SILConverters
Client applications
FieldWorks (as of v3.1)
Clipboard EncConverter
Data Conversion Macro.dot (MS Word)
Bulk Word Document Converter
Adapt It (as of v3.1)
Bulk SFM Data Converter
SpellFixer.dot (MS Word)
SILConverters for Office (Access, Excel, Publisher)
TECkit map Unicode Editor
XML Document Data Converter
Consistency Spelling Checker.dot (MS Word)
Discourse Chart Builder
EncConverters core
EncConverters Repository API
Transduction Engines
TECkit (compiled .TEC) (compilable .MAP)
Compound Converter (series)
Windows code page converters
Adapt It Knowledge Base Lookup
Consistent Changes (CC)
Primary-Fallback Converter (parallel)
Python script functions
Adapt It Target Word Guesser (esp. for use in Adapt It)
ICU
• Transliterators
• Converters
• Regular Expression Find and Replace
Perl Expressions

Figure 1: SILConverters Suite

Installation Features

This document gives more detail about the boxes in Figure 1 and refers you to further information about how use the different utilities and applications for different text transduction applications. The information in this document is organized around the different components available from the *Feature Selection Tree* in the SIL Converters Installer.

Feature overview


SIL Encoding Converters 2.6 Setup
Select Features
Please select which features you would like to install.
SILConverters' client applications
Transduction Engines
Maps and Tables
Additional TECkit applications
Wise Installation Wizard (R)
Disk Cost
Reset
< Back
Next >
Cancel

Feature Description:
Expand this node to select or deselect individual client applications

This feature will be installed on the local hard drive.

This feature requires 0KB on your hard drive. It has 5 of 7 subfeatures selected. The subfeatures require 28MB on your hard drive.

Figure 2: Feature Selection Tree

As you can see from Figure 2, there are four main categories of features that you can choose from when installing SIL Converters:

SIL Converters' client application

This feature node contains some of the programs at the top layer of Figure 1, which are generally of the most interest to end users. These programs and utilities allow you to convert text data (e.g. Word documents, SFM documents, XML Documents, data on the system clipboard) using the text processing capabilities provided by the various transduction engines at the bottom of Figure 1.

Transduction Engines

This feature node contains the different transduction engine components that provide text processing capabilities at the lowest layer of Figure 1.

Most users should accept the defaults for this feature to insure that the proper transduction engines are installed. Otherwise, you must make sure you have the required transduction engines installed for the different text processing tasks you want to do.

Examples:

- If you intend to do encoding conversions, you probably need to install the TECkit and/or Consistent Changes (CC) transduction engines.
- If you want to use an ICU transliterator, you need to install the IBM Components for Unicode transduction engine.
- If you want to use an Adapt It knowledge base to provide lookup capabilities (Source to Target word) or to use the new target word guesser transducer in Adapt It, then you need to install the Adapt It transduction engines.
- If you want to write Perl expressions or call Python script functions for text processing, you need to install one or both of those transduction engines (both of which require separately available program distributions—see below)

Maps and Tables

This feature node has packages containing individual instances of conversion maps and tables (e.g. for TECkit and/or CC) grouped together logically.

A few of the items are useful for all users, such as the *Basic Converters and ICU Transliterators* sets. Otherwise, you can only install those converter sets you expect to need (e.g. based on your entity).

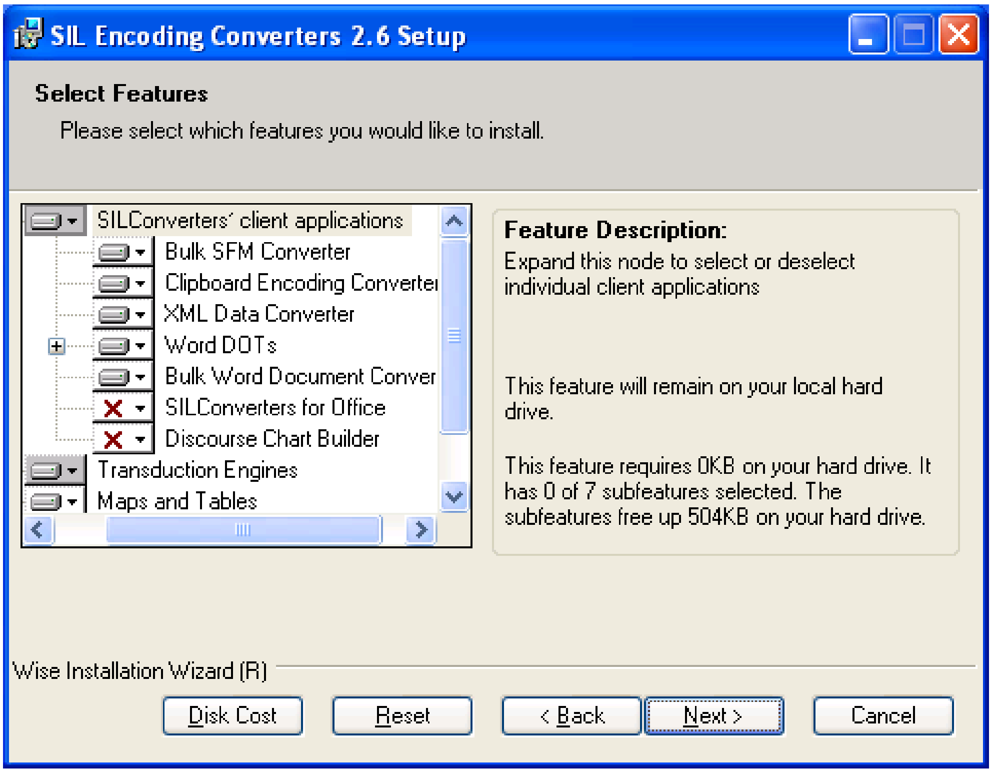
If you would like to add a package of converters to the SIL Converters' installer, contact ✉ mailto:silconverters_support@sil.org.

Additional TECkit applications

Since the SILConverters installer installs TECkit (a subfeature of the Transduction Engine feature node discussed above), this feature node adds the rest of the content of the TECkit download from the TECkit site (i.e. the documentation and other TECkit client applications described at http://scripts.sil.org/TECkitIntro). A new *TECkit map Unicode Editor* (not part of the TECkit download) assists in the creation of TECkit maps available from this feature node and is discussed below.

The following sections describe the sub-features available in each of these four nodes.

SIL Converters’ client applications



SIL Encoding Converters 2.6 Setup

Select Features

Please select which features you would like to install.

Feature Description:
Expand this node to select or deselect individual client applications

This feature will remain on your local hard drive.

This feature requires 0KB on your hard drive. It has 0 of 7 subfeatures selected. The subfeatures free up 504KB on your hard drive.

Wise Installation Wizard (R)

Figure 3: SILConverters’ Client Applications

The SILConverters’ installer installs the following of SIL Converters client applications (see Figure 1).

- Bulk SFM Converter
- Bulk Word Document Converter
- Clipboard Encoding Converter
- SILConverters for Office
- XML Data Converter
- MS Word Document Template Converters
- Discourse Chart Builder

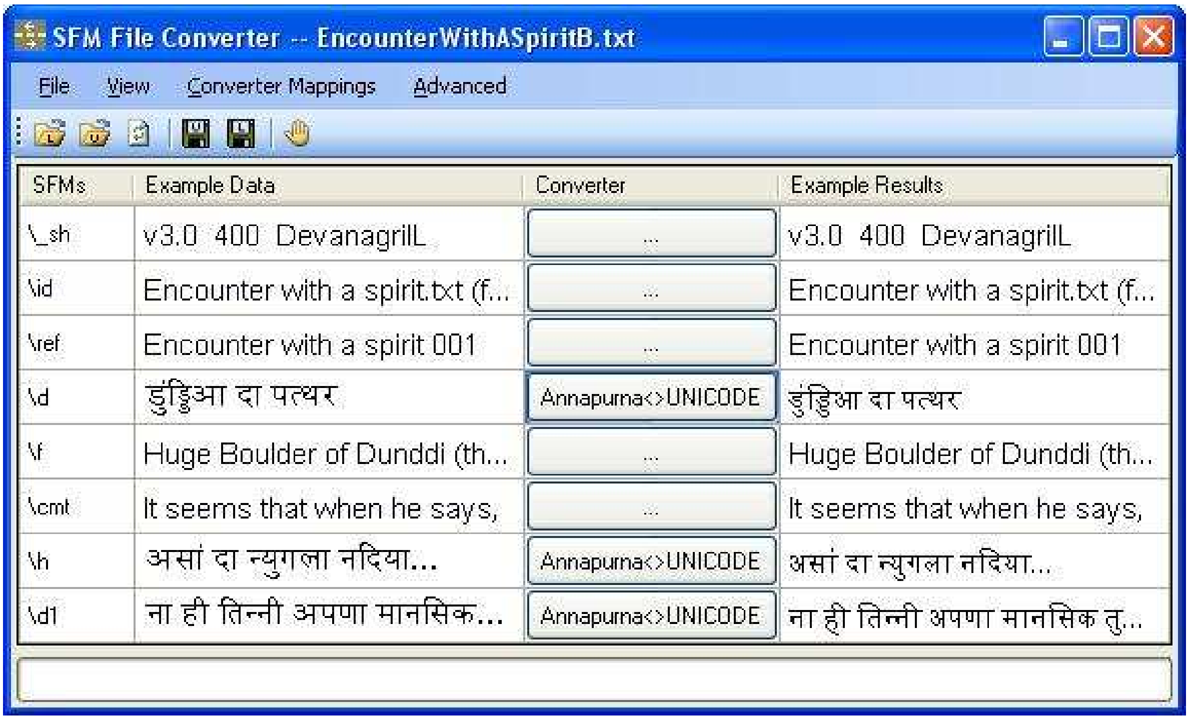
The *FieldWorks* and *Adapt It* client applications have separate install programs.

Bulk SFM Converter

Use this application to convert the data in Standard Format Marker (SFM) fields using converters from the EncConverters’ system repository and to convert the encoding of data in Shoebox, Toolbox, and Paratext (SFM) documents. You can also open multiple SFM documents for processing at the same time.

To run the program

- Click **Start… / All Programs / SIL Converters / Bulk SFM Converter**.
- For help, click in the main window area and press the F1 key.



SFMs	Example Data	Converter	Example Results
_sh	v3.0 400 DevanagriIL	...	v3.0 400 DevanagriIL
\id	Encounter with a spirit.txt (f...	...	Encounter with a spirit.txt (f...
\ref	Encounter with a spirit 001	...	Encounter with a spirit 001
\d	डुंड्डिआ दा पत्थर	Annapurna<>UNICODE	डुंड्डिआ दा पत्थर
\f	Huge Boulder of Dunddi (th...	...	Huge Boulder of Dunddi (th...
\cmt	It seems that when he says,	...	It seems that when he says,
\h	असां दा न्युगला नदिया...	Annapurna<>UNICODE	असां दा न्युगला नदिया...
\d1	ना ही तिन्नी अपणा मानसिक...	Annapurna<>UNICODE	ना ही तिन्नी अपणा मानसिक तु...

Figure 4: Bulk SFM File Converter

To use

The Bulk SFM File Converter can be used as follows:

1. To open one or more SFM files for conversion, click the **File / Open SFM Documents** menu item. This brings up two sub-menus that you choose from depending on the encoding of the files. If you are converting a non-Unicode (i.e. Legacy-encoded) Shoebox project, for example, then choose the **Non-Unicode** menu. If the data in the SFM files is already Unicode-encoded, then choose **Unicode**.

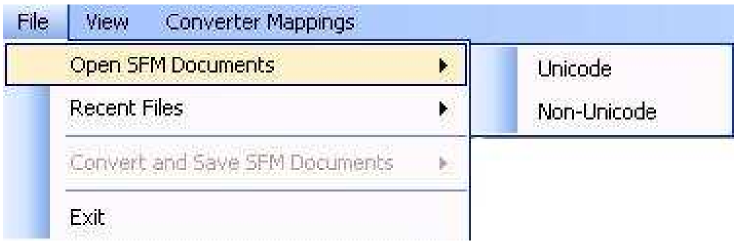


Figure 5: Bulk SFM File Converter Open menu commands

> Note that all documents opened at the same time should be the same encoding, though they don't all have to have the same list of SFM fields. However, the same SFM field in different files should have the same meaning (i.e. contain data encoded with the same font).

- You can also use the toolbar buttons 📂 and 📂 to launch the **File / Open** window for Legacy (Non-Unicode) and Unicode-encoded files respectively.
- The program will scan all the selected files and list all of the unique SFM fields in column 1 of the table in the center portion of the program window.

2. For each SFM field (row of the table), you will see some sample data in the **Example Data** column.
    - You can click on a cell in the **Example Data** column to see more data of the same SF marker.
    - You can also click with the right mouse button in order to configure the font to be used to display the data in the cells of the **Example Data** column.
3. Click on the buttons in the **Converter** column to display the **Choose Converter** window. This allows you to select a converter from the system repository to use for converting the text in that row (i.e. for the corresponding SFM field in column 1).
    - You can repeat the selection of a converter for other rows by clicking on the **Converter** box of another row with the right-mouse button. The last converter chosen will be repeated for the new row.
    - You can remove the converter mapping for a given row by clicking on the button again to launch the **Choose Converter** dialog and then click the Cancel button. This will reset the converter mapping for that row.
4. Once you've selected a converter for a particular row, the **Example Results** column will preview what the data looks like after the conversion.
    - As with the **Example Data** column, you can click with the right mouse button in the **Example Results** column in order to configure the font to be used to display the data in the cells of that column.
5. Once you have configured the converters to apply, you are ready to do the conversion. However, if you want to save these settings (i.e. the converter and any configured display fonts to use for each SFM field), you can use the commands in the **Converter Mappings** menu:

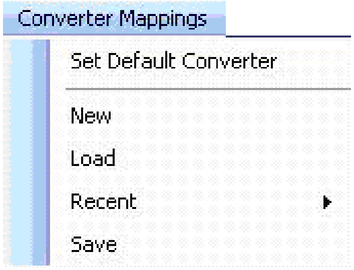


Figure 6: Converter Mapping menu commands

- **Set Default Converter:** allows you to select a converter to be applied to all rows in the table that aren't currently configured.

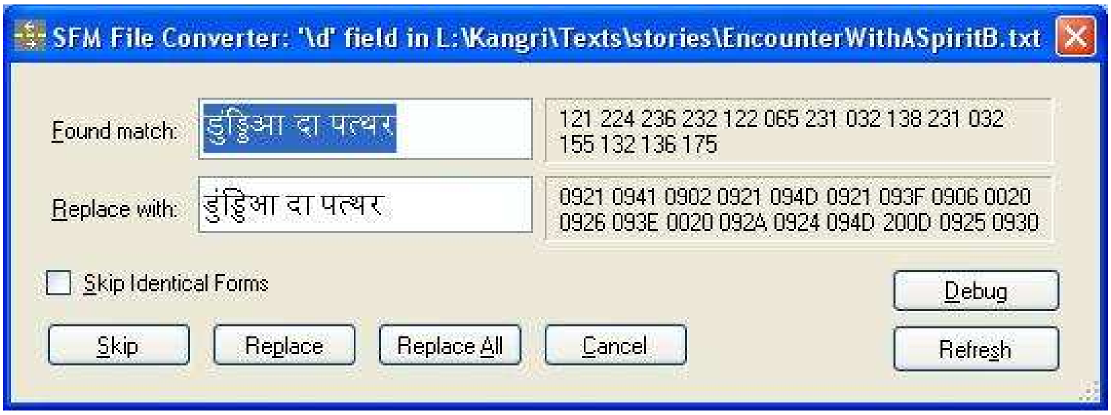
- **New:** resets all the converter and font mappings.
- **Load:** loads a previously saved converter and font mapping set.
- **Recent:** shows previously saved converter and font mapping sets that you can click on to load.
- **Save:** saves the currently configured converter and font mapping set in a file you choose.

6. To initiate the conversion, click the **File / Convert and Save Document** command or click the corresponding toolbar buttons: (save as Unicode-UTF8) and (save as Non-Unicode/Legacy-encoding). The program will then convert all of the text in the file(s) with the converters that you've chosen. When the conversion of the first document is complete, a Save dialog will appear allowing you to save the document with a new name or in a new folder.

> It is highly recommended that you do not overwrite the original document unless you are sure that you have an adequate backup of the file.

7. There are two additional features to be aware of:

   - You can use the **File / Reload** menu command (or the corresponding toolbar button) to reload the original document(s) if the conversion is stopped for any reason or you want to restart it from the beginning.
   - You can turn on the **Advanced / Single-step conversion** menu item (or the corresponding toolbar button) to execute the conversions one run of text at a time. In this mode, you will see a dialog which shows the result of the conversion for each field of data in the document. The following image shows an example:



SFM File Converter: '\d' field in L:\Kangri\Texts\stories\EncounterWithASpiritB.txt

Found match: डुंड्डिआ दा पत्‍थर

121 224 236 232 122 065 231 032 138 231 032 155 132 136 175

Replace with: डुंड्डिआ दा पत्‍थर

0921 0941 0902 0921 094D 0921 093F 0906 0020 0926 093E 0020 092A 0924 094D 200D 0925 0930

□ Skip Identical Forms

Figure 7: Advanced Single-step Conversion mode dialog

   - The **Found match** box shows the legacy data and the **Replace with** box shows the results after the conversion to Unicode. This mode is useful if you aren't quite sure whether the mapping file is working correctly.

The *SILConverters for Office* application has a feature that allows you to test the the round-trip capability of your conversion map in Microsoft Word. If you want to test your converter map, see the description of the feature below.

Bulk Word Document Converter

Use this application to convert one or more Word documents using converters from the system repository. Though it has similar functionality to the Data Conversion Macro, this tool is used outside of the Microsoft Word environment and can be used on multiple documents at the same time. It is also recommended for documents that have very complex formatting, which the Data Conversion Macro occasionally has difficulty converting correctly. There is also the capability to automatically search your hard-drive for documents that use a specific font or fonts through which to do the data conversion.

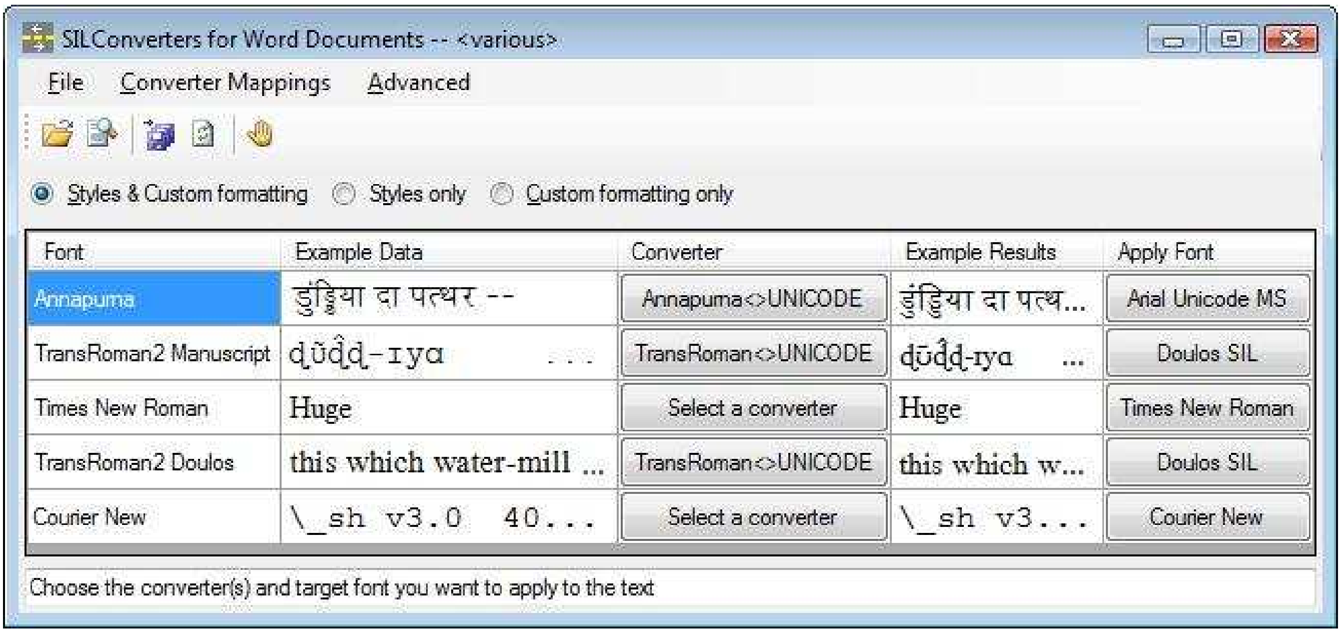
This application requires the *Microsoft Word Primary Interop Assemblies* (PIAs) to be installed. These can only be installed either from your Microsoft Office/Word Professional 2003 CD/DVD or from a 4MB download available from the Microsoft website and only if you have Microsoft Word 2003 Professional installed on your machine.

On some machines, these assemblies will automatically be installed the first time they are requested, so you can try the program first to see if this is the case for your computer. However, if the program crashes, see the FAQ page at http://msdn2.microsoft.com/en-us/library/aa159923(office.11).aspx, for details on installing the Office Primary Interop Assemblies via one of the other two methods.

This program will also work if you have Word 2007 installed on your machine (i.e. for *.docx files), however, it currently cannot process native Word 2007 documents without first converting them to use Word 2003 compatibility mode. So if you attempt to use this tool to convert Word 2007 *.docx files, some formatting may be lost.

To run program

- Click **Start… / All Programs / SIL Converters / Bulk Word Document Converter**.



● Styles & Custom formatting ○ Styles only ○ Custom formatting only

Font	Example Data	Converter	Example Results	Apply Font
Annapurna	इड्डिया दा पत्थर --	Annapurna<>UNICODE	इंड्डिया दा पत्थ...	Arial Unicode MS
TransRoman2 Manuscript	ḍũḍ̂ḍ-ɪyɑ . . .	TransRoman<>UNICODE	ḍũḍ̂ḍ-ɪyɑ ...	Doulos SIL
Times New Roman	Huge	Select a converter	Huge	Times New Roman
TransRoman2 Doulos	this which water-mill ...	TransRoman<>UNICODE	this which w...	Doulos SIL
Courier New	_sh v3.0 40...	Select a converter	_sh v3...	Courier New

Choose the converter(s) and target font you want to apply to the text

Figure 8: Bulk Word Document Converter

This tool allows you to specify different converters to use to convert each font in the document(s) that are loaded. You can also choose a different font to be applied to the converted text by select a new font name in the Apply Font column.

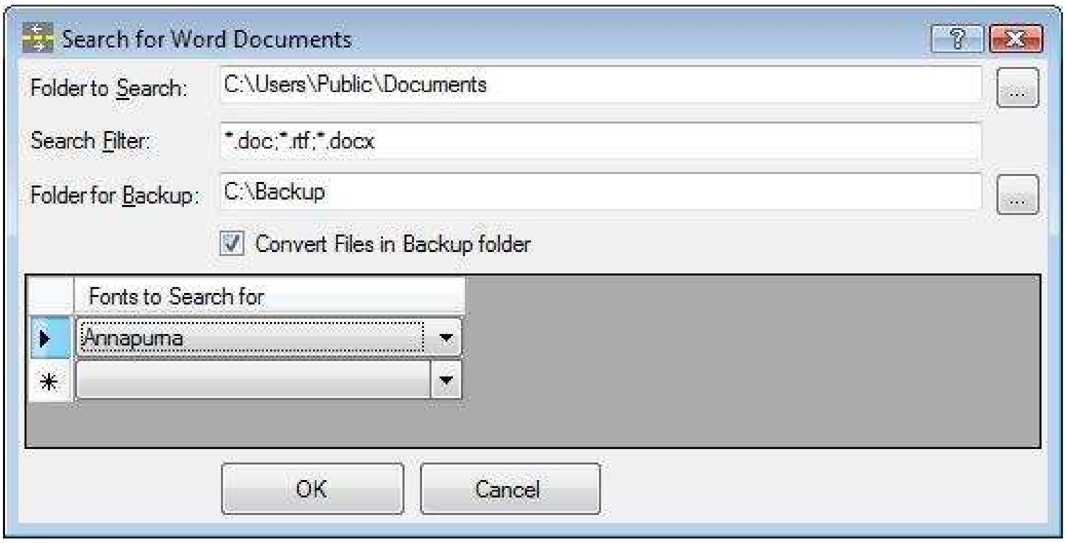
To use

The Bulk Word Document Converter can be used as follows:

1. Open one or more Word documents using the **File / Open** command (or the toolbar button), or with the **File / Search** command (or the toolbar button)

> Note that you can open any file type that Microsoft Word is capable of opening (e.g. DOC, RTF, XML, HTML, etc). Text files (e.g. SFM documents), however, are not recommended because they typically won't have any font/formatting information associated with the encoded text.
>
> If you want to convert an SFM document, see the *Bulk SFM Document Converter* above.
>
> Also, in order to process Word 2007 documents (i.e. *.docx), you must either have Word 2007 installed or the *Microsoft Office Compatibility Pack for Word, Excel, and PowerPoint 2007 File Formats* (available from the webpage: http://www.microsoft.com/downloads/details.aspx?FamilyID=941b3470-3ae9-4aee-8f43-c6bb74cd1466&displaylang=en), as well as *Word 2003*.

- If you use the **File / Search** command, the following dialog box will be displayed:


Search for Word Documents

Folder to Search: C:\Users\Public\Documents

Search Filter: *.doc;*.rtf;*.docx

Folder for Backup: C:\Backup

☑ Convert Files in Backup folder

Fonts to Search for
Annapurna

Figure 9b: Search for Word Documents window

- In this dialog, you can specify the root folder to begin the search (e.g. here, C:\Users\Public\Documents), the search filter, the root folder where the files to be converted will be backed up (here, C:\Backup). You can also check the **Convert Files in Backup folder** box to have the documents that were copied to the Backup folder be converted rather than the documents in their original location. Finally, you can also enter one or more fonts to search for. Only documents which have one or more of the listed fonts will be copied to the backup location and opened for conversion.

- Once you open the file(s), the program will scan all of the selected documents for both Style-based and Custom-formatted text. Then it will list all of the unique fonts encountered in the table at the center of the program window.

- Normally, it is best to use the default ● Styles & Custom formatting, button to convert your text since this should convert everything. However, depending on your document, you can also choose one of the other two buttons to have the program convert ○ Styles only, or ○ Custom formatting only data.

2. For each font listed in the **Font** column, you will see some example data in the **Example Data** column.

    - You can click on a cell in the second (Example Data) column to see the next run of text in the same font.

3. Clicking on the buttons in the **Converter** column displays the **Select Converter** window which allows you to choose a converter from the system repository to use for converting the text in that row (i.e. corresponding to the font in column 1).

    - You can repeat the selection of a converter for other rows by clicking on the **Converter** box of another row with the right-mouse button. The last converter chosen will be repeated for the new row.

- You can remove the converter mapping for a row by clicking on the button again to launch the **Select Converter** dialog and then click the Cancel button. This will reset the converter mapping for that row.
- Once you've selected a converter for a particular row, the **Example Results** column will preview what the data looks like after the conversion.

4. Clicking on the buttons in the Apply Font column allows you to choose which font to apply to the text after conversion.

> It doesn't matter what size you choose for the font in the **Font Selection** dialog box, because this program will not change the font size of the text in the document. If you need to adjust the size of the font of the converted text, then you have to do it either after conversion in Word itself or using the Data Conversion Macro instead which supports increasing the size of the applied target font (e.g. by some percentage).

5. Once you have configured all of the converters and fonts to apply in the third and fifth columns of the table, then you are ready to do the conversion. However, if you want to save these settings (i.e. the converter to use and the font to apply for each row in the table), you can use the commands in the **Converter Mappings** menu:

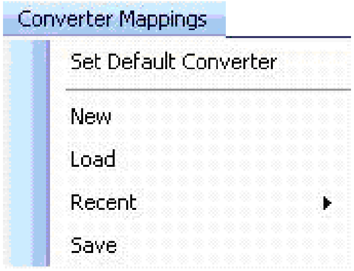


Figure 10: Converter Mapping menu commands

- **Set Default Converter:** allows you to select a converter to be applied to all rows in the table that aren't currently configured.
- **New:** resets all the converter and font mappings.
- **Load:** loads a previously saved converter and font mapping set.
- **Recent:** shows previously saved converter and font mapping sets that you can click on to load.
- **Save:** saves the currently configured converter and font mapping set in a file you choose.

6. To initiate the conversion, click the **File / Convert and Save Documents** command or click the corresponding toolbar button: . The program will then convert all of the text in the file(s) using the selected converters and fonts to apply that you've chosen. When the conversion of the first document is complete, a Save dialog will appear allowing you to save the document with a new name or in a new folder.

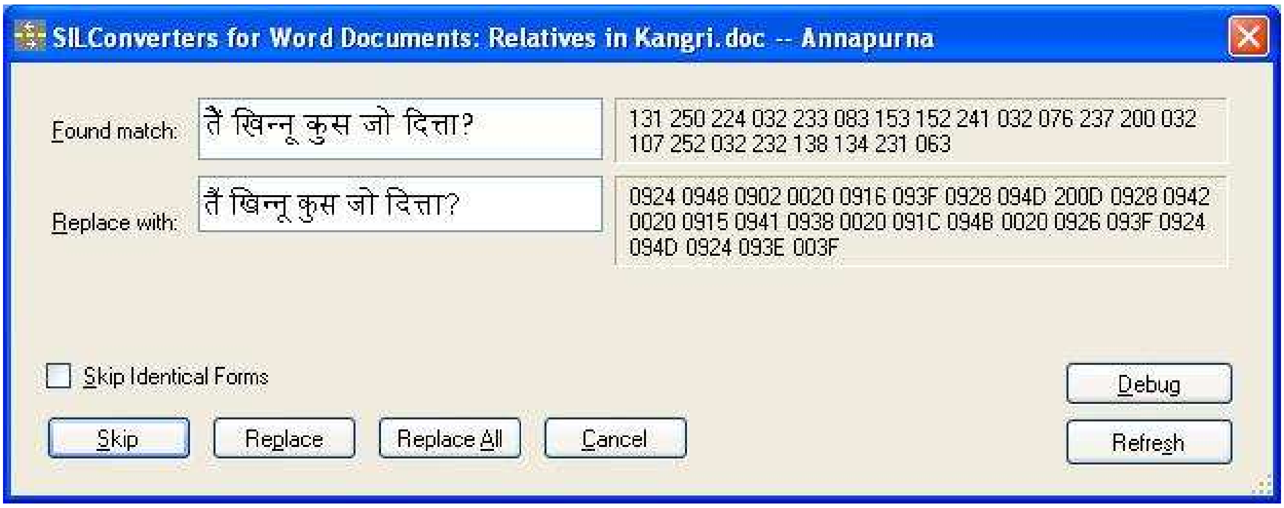
> It is highly recommended that you do not overwrite the original document unless you are sure that you have an adequate backup of the file.

> If you loaded the files as a result of the **File / Search** command, the files will automatically be saved either as the original filename in the original folder location (after a backup has been made to the specified Backup Folder), or the files copied to the Backup Folder will be converted and automatically saved after conversion. That is, you do not need to give the names by which the files will be saved in this mode.

7. There are two additional features to be aware of:

    - You can use the **File / Reload** menu command (or the corresponding toolbar button) to reload the original document(s) if the conversion is stopped for any reason and you want to restart it from the beginning.

    - You can turn on the **Advanced / Single-step conversion** menu item (or the corresponding toolbar button) to execute the conversions one run of text at a time. In this mode, you will see a dialog which shows the result of the conversion for each run of text in the document. The following image shows an example:



SILConverters for Word Documents: Relatives in Kangri.doc -- Annapurna

Found match: तैं खिन्नू कुस जो दित्ता?

131 250 224 032 233 083 153 152 241 032 076 237 200 032 107 252 032 232 138 134 231 063

Replace with: तैं खिन्‍नू कुस जो दित्ता?

0924 0948 0902 0020 0916 093F 0928 094D 200D 0928 0942 0020 0915 0941 0938 0020 091C 094B 0020 0926 093F 0924 094D 0924 093E 003F

□ Skip Identical Forms

Figure 11: Advanced Single-Step Conversion mode dialog

- The **Found match** box shows the legacy data and the **Replace with** box shows the results after the conversion to Unicode. This mode is useful if you aren't quite sure whether the mapping file is working correctly.

> The *SILConverters for Office* application has a feature that allows you to test the the round-trip capability of your conversion map in Microsoft Word. If you want to test your converter map, see the description of the feature below.

Clipboard Encoding Converter

This utility can be used to convert text copied to the Windows clipboard.

To start it up

The Clipboard Encoding Converter is a untilty that you access from the Windows Task Bar. To use it you need to first start it up.

- To start it only when you need it: Click **Start… / All Programs / SIL Converters / Clipboard EncConverter**.
- To start it every time you start Windows: Add a shortcut to *Clipboard EncConverter* to your **All Programs / Startup folder**.

To use

If you click on the Clipboard Encoding Converter icon with the right mouse button, the following popup menu appears:

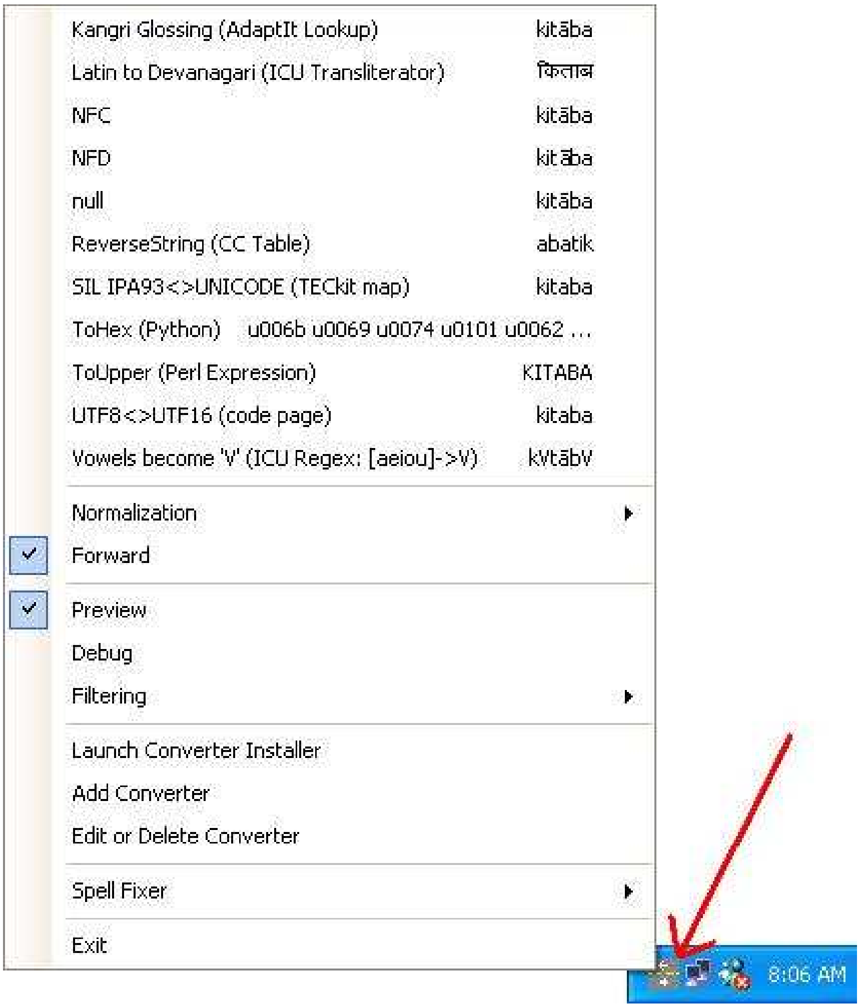

Kangri Glossing (AdaptIt Lookup) kitāba
Latin to Devanagari (ICU Transliterator) किताब
NFC kitāba
NFD kitāba
null kitāba
ReverseString (CC Table) abatik
SIL IPA93<>UNICODE (TECkit map) kitaba
ToHex (Python) u006b u0069 u0074 u0101 u0062 …
ToUpper (Perl Expression) KITABA
UTF8<>UTF16 (code page) kitaba
Vowels become 'V' (ICU Regex: [aeiou]->V) kVtābV
Normalization
✓ Forward
✓ Preview
Debug
Filtering
Launch Converter Installer
Add Converter
Edit or Delete Converter
Spell Fixer
Exit
8:06 AM

Figure 12: EncConverter Clipboard Mode popup

To convert text

1. Copy some text by selecting it and pressing Ctrl + C (for example, in Microsoft Word).
2. *Right-click* the Clipboard EncConverter Icon in the Windows task bar (*System Tray*—see red arrow in Figure 12).

The available conversions will be listed in the top section of the popup window. If **Preview** menu is checked, a sample of your copied text will appear as it will be converted to the right of each conversion name.

3. In the popup menu, select the conversion you want.
4. Paste your converted text where you want it by pressing Ctrl + V.

Menu options

This section explains the different menu options available on the Clipboard EncConverter popup menu:

- **Normalization:** this menu setting takes the values, None (default), Decomposed, or Composed. This can be used to request a particular Unicode Normalization Form for the output of the conversion. This only applies when the output of the conversion is Unicode-encoded.
- **Forward:** this menu setting is either checked or uncheck. If checked (default), then the requested conversion will be executed in the forward direction. If you want the conversion to occur in reverse direction (and your converter supports bidirectional conversions), then uncheck this menu item.
- **Preview:** this menu setting is either checked or unchecked. If checked (default), then the current text contents of the clipboard will be converted using each of the available converters and the results will be shown to the right of the converter name in the popup menu. If unchecked, then the results of the conversion will not be previewed in the popup menu.
- **Debug:** this menu setting is either checked or unchecked. If checked, then diagnostic messages will be displayed that show the data sent and received by the different parts of the EncConverters utility. This can be helpful to debug whether a mapping converter is operating correctly or not.
- **Filtering:** this menu setting allows you to filter the converters displayed in the top portion of the popup menu. You can filter converters based on the transduction type (e.g. Unicode Encoding Converters or Transliterators), implementation type (e.g. CC table or TECkit map), or based on the technical encoding name (e.g. UNICODE, SIL-IPA93-2001).
- **Launch Converter Installer:** this menu command will bring up the **Converter Installer** program from which you can add installed converters to your system repository so they become available to different client programs (see *Adding Converters to the System Repository*).
- **Add Converter:** this menu command brings up the **Choose a Transduction Engine** window from which you can add an existing converter (e.g. a TECkit map located on your machine but not already in the system repository) or create a new one (e.g. a new Perl Expression). See *Choose a Transduction Engine dialog box.*
- **Edit or Delete Converter:** this menu command brings up the **converter selection** dialog box from which you can edit or delete a converter in the system repository.

- **SpellFixer:** this menu brings up the different SpellFixer-related commands that are available as shortcuts from the **Clipboard EncConverter** popup menu. See the SpellFixer.dot file for details of these commands (Click **Start… / All Programs / SIL Converters / Help / Help for the SpellFixer macro**).
- **Exit:** this menu command will cause the **Clipboard EncConverter** program to stop running and remove itself from the system tray.

SILConverters for Office

This Office add-in gives SILConverters support in Microsoft Publisher, Excel, Access, and Word.

> It is generally not recommended to use this add-in for data conversion in Microsoft Word. Instead, for conversion of Word documents, it is recommended that you use either the Bulk Word Document converter or the Data Conversion Macro. The SILConverters for Office add-in however adds a few useful commands to Microsoft Word related to Regular Expression searching and bidirectional conversion (e.g. TECkit) map validation. See the section below on these new features.

> This application requires the Office Primary Interop Assemblies (PIAs) to be installed. These can only be installed either from your Microsoft Office Professional CD/DVD or from a 4MB download available from the Microsoft website and only if you have Microsoft Office Professional (either 2003 or 2007) installed on your machine.
>
> On some machines, these assemblies will automatically be installed the first time they are requested, so you can try the program first to see if this is the case for your computer. However, if the program crashes, see the FAQ page at http://msdn2.microsoft.com/en-us/library/aa159923(office.11).aspx, for details on installing the Office Primary Interop Assemblies via one of the other two methods.
>
> It is also required that you have all four products that can use the *SILConverters for Office* assembly installed: *Word*, *Excel*, *Publisher*, and *Access*. If any of these programs are not installed (either 2003 or 2007 versions), then *SILConverters for Office* will not show up in the feature tree during installation.

> If you install the SILConverters for Office application in one user account, it may not automatically be available to Office applications in another user account. To enable it in another account, you have to use the "COM Add-ins" dialog in the respective program to browse for and load the SILConvertersOfficeShim.dll assembly, which is installed in the C:\Program Files\SIL\SILConverters folder by default.

To use in Microsoft Publisher:

The SILConverters for Office add-in is configured to automatically load when Microsoft Publisher starts. However, depending on your Macro/Add-in security settings, you may need to manually enable it (see instructions below). You can tell whether this is needed or not based on whether the SILConverters menu (circled in red below in the image below) is visible or not.

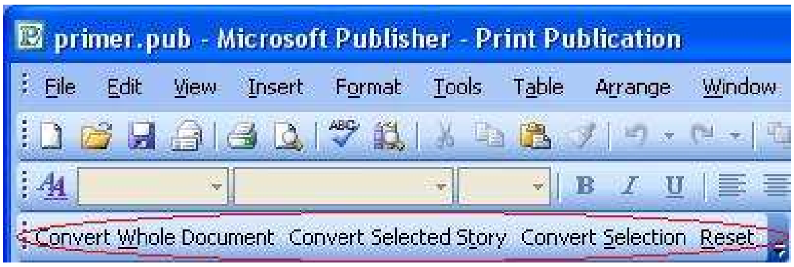
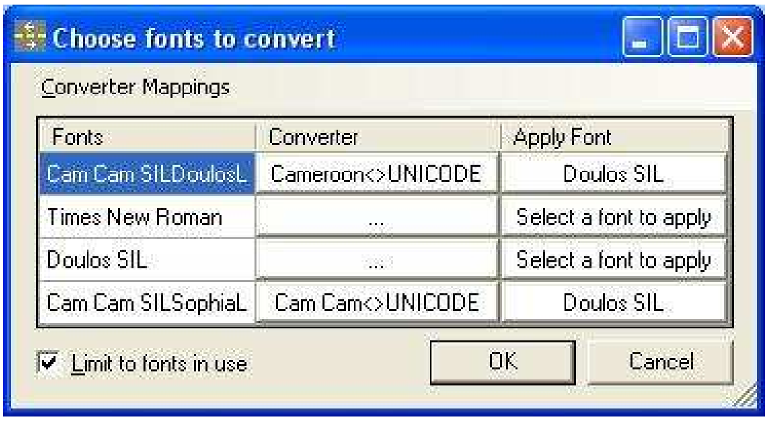


Figure 13: SILConverters menu in Microsoft Publisher

> There is not a significant difference in terms of this menu between Publisher 2003 and Publisher 2007.

Once the add-in is enabled, there are three commands on the menu bar you can use to convert the text in your Publisher document. The commands are: **Convert Whole Document**, **Convert Selected Story**, and **Convert Selection**.

- **Convert Whole Document:** this menu command can be used to convert all of the Stories (i.e. the text in connected text frames) in the open document. This command will first launch the font/converter mapping dialog through which you can specify which converter to use to convert the different fonts in the document:



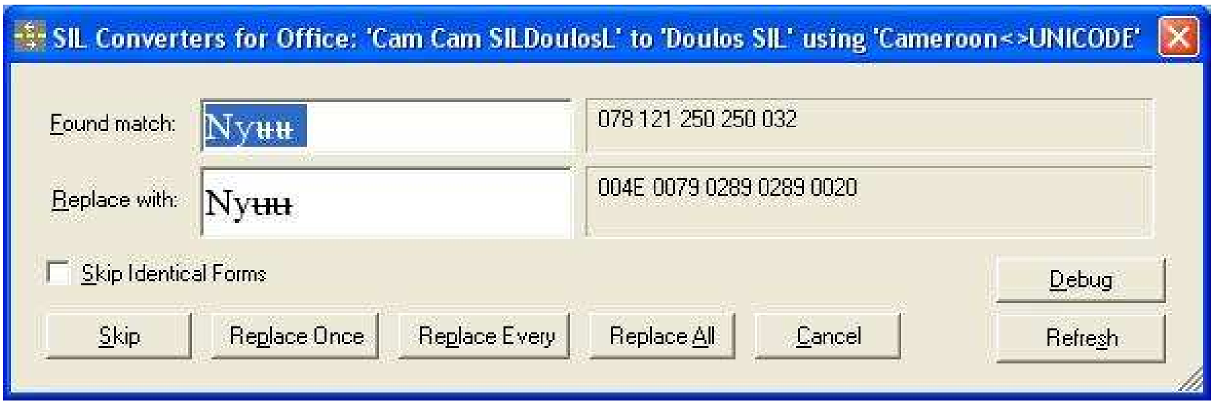
Fonts	Converter	Apply Font
Cam Cam SILDoulosL	Cameroon<>UNICODE	Doulos SIL
Times New Roman	...	Select a font to apply
Doulos SIL	...	Select a font to apply
Cam Cam SILSophiaL	Cam Cam<>UNICODE	Doulos SIL

☑ Limit to fonts in use

Figure 14: Converter and Target Font selection dialog in Microsoft Publisher

As with other SILConverter client applications, this dialog also supports saving and loading the mapping information in files via the commands in the **Converter Mappings** menu.

Once you click OK , the add-in will go through all of the Text Frames in the document and convert the text according to the information you specified in the **Converter Mappings** dialog. For each word in the document, the following dialog will be displayed:



SIL Converters for Office: 'Cam Cam SILDoulosL' to 'Doulos SIL' using 'Cameroon<>UNICODE'

Found match: Nyʉʉ 078 121 250 250 032

Replace with: Nyʉʉ 004E 0079 0289 0289 0020

□ Skip Identical Forms

Figure 15: Conversion result dialog

This dialog box shows you the original word (in the **Found Match** box) and the converted word (in the **Replace With** box). The code point values of the text in those two boxes are shown in the two boxes to the right.

If the conversion is correct, you can click on one of the **Replace** buttons to have the converted value replace the original value in the document.

The other buttons have the following functions:

- **Skip:** this will leave the word as is in the original document (i.e. no conversion).
- **Replace Once:** This will replace the original word with the converted word, but only for this instance of the original word.
- **Replace Every:** This will replace the original word with the converted word, and if the same word occurs again, it will be automatically replaced.
- **Replace All:** This will cause all of the words in the current Text Frame to be converted and replaced.
- **Cancel:** This will stop the conversion process and dismiss the dialog box.
- **Debug:** This will repeat the conversion between the Found Match text and the Replace With text showing informational dialogs about what values were received by different processes during the conversion process.
- **Refresh:** If you change the underlying converter map and re-compile, this button will redo the conversion.

- **Convert Selected Story:** this menu command is similar to the **Convert Whole Document** command except that it will convert all of the text in the currently selected story only.
- **Convert Selection:** this menu command is similar to the other two commands except that it will only convert the selected text.
- **Reset:** this command will reset the state of all the commands executed by the add-in. You might need to use this to clear conversion information between conversions (e.g. what converters to use, where a conversion was left off, etc).

Loading the Add-in

If the SILConverters' menu is not visible, it either means that your security settings are preventing add-ins from loading or that you are in a different Windows user account than when the SILConverters for Office feature was installed.

To enable the add-in in *Publisher,* do the following:

1. Click **Tools / Macro / Security...**

   In *Publisher 2003*, the **Security** dialog will be displayed. To enable the add-in, you can change your Security Level setting to Medium and restart Microsoft Publisher. When it restarts, the **Security Warning** dialog will be displayed from which you can request to **Enable Macros**.

   Another way to enable the add-in is to switch to the **Trusted Publishers** tab in the **Security** dialog and check the box that says, **Trust all installed add-ins and templates** and restart Microsoft Publisher. When it restarts, the add-in should be enabled.

   In *Publisher 2007*, the Trust Center dialog box appears.

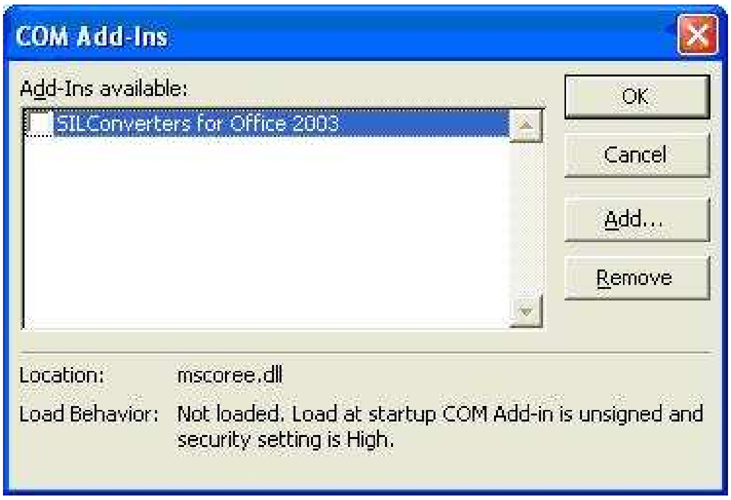
   Click the **Add-ins** item in the left-hand pane and make sure the **Disable all Application Add-ins** box is not checked.

2. Click **Tools / Macro / Security...**

   The **Security** dialog will be displayed. To enable the add-in, you can change your Security Level setting to Medium and restart Microsoft Publisher. When it restarts, the **Security Warning** dialog will be displayed from which you can request to **Enable Macros**.

   Another way to enable the add-in is to switch to the **Trusted Publishers** tab in the **Security** dialog and check the box that says, **Trust all installed add-ins and templates** and restart Microsoft Publisher. When it restarts, the add-in should be enabled.

3. If this doesn't cause the SILConverters menu to be displayed, you should click **Tools / Add-Ins** (or for Publisher 2007, in the **Manage** area of the Trust Center dialog, select **COM Add-ins** and then Go) to see the error that is preventing it from loading. For example:



COM Add-Ins

Add-Ins available:

□ SILConverters for Office 2003

Location: mscoree.dll

Load Behavior: Not loaded. Load at startup COM Add-in is unsigned and security setting is High.

Figure 16: COM Add-Ins dialog in Microsoft Publisher

If the **SILConverters for Office** item is not in the COM Add-Ins list, then click the **Add** button to browse for and load the file: C:\Program Files\SIL\SILConverters\SILConvertersOfficeShim.dll.

If the checkbox next to the "SILConverters for Office" add-in in the COM Add-ins dialog is checked and the Load Behavior message below says "Load on Demand, (not currently loaded)", you can load it by first unloading it and then loading it again. If you want to have the add-in load when Word initially starts, change the value of the following registry key to **3**: HKEY_CURRENT_USER\Software\Microsoft\Office\Publisher\Addins\SILConvertersOffice.Connect [LoadBehavior]

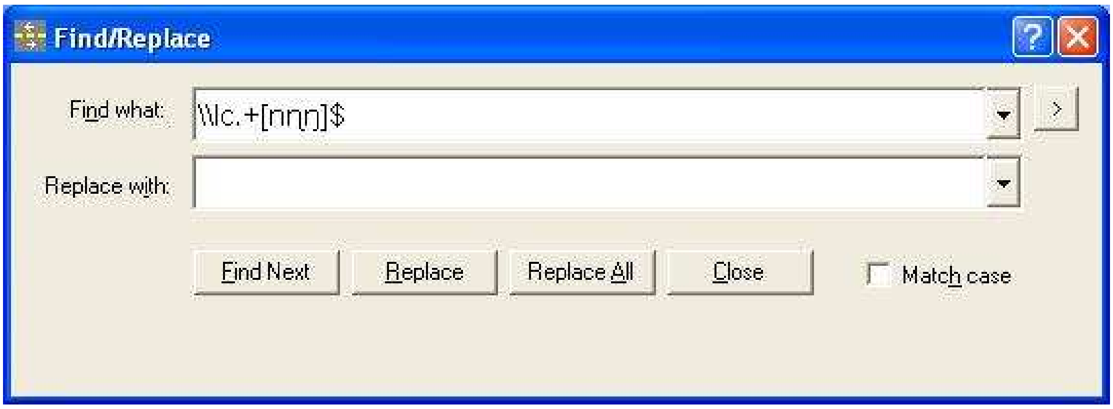
If this doesn't solve the problem, insure that the following updates are installed on your machine: Control Panel, Add/Remove Programs, **Shared Add-in Extensibility Update for Microsoft .NET Framework 2.0 (KB908002)** and **Shared Add-in Support Update for Microsoft .NET Framework 2.0 (KB908002)**. If these are not presently installed (which can happen if you don't use the master installer to install SILConverters), then run the following program to install them: C:\Program Files\SIL\SILConverters\MS KB908002 Fix\setup.exe.

To use in Microsoft Word:

To convert the encoding of text in Microsoft Word documents, it is recommended that you use either the **Bulk Word Document converter** or the **Data Conversion Macro**. The **SILConverters for Office** add-in does work in Microsoft Word, but it provides features related to Regular Expression searching and bidirectional conversion (e.g. TECkit) map validation rather than specifically data conversion.

The SILConverters menu provides 4 menu commands in Microsoft Word.

- **Convert Selection:** this menu command will go through the selected text word-by-word and for each font encountered, it will query for a converter to use and a font to apply. Then it will display the **Conversion result** dialog to allow you to confirm the replacement. See the section above on Microsoft Publisher for details of that dialog and the function of the different buttons.
- **Regex Find/Replace:** this command provides the ability to use ICU Regular Expressions to search the text in a Word document. When you click this menu, the following **Find/Replace** dialog box is displayed:



Find what: \\lc.+[nɳŋ]$

Replace with:

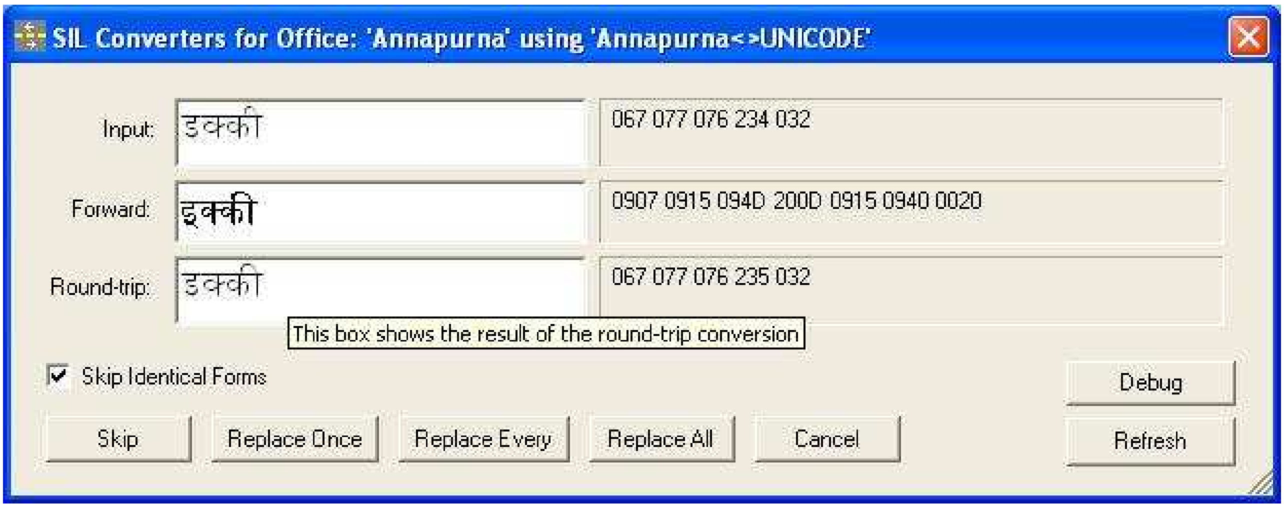
□ Match case

Figure 17: Regular Expression Find/Replace dialog in Microsoft Word

- The expression in the **Find what** box says to find any text that begins with the string \lc followed by one or more characters (.+) until it finds one of the three nasal characters inside the pattern matching brackets “[” “]” (i.e. nɳŋ) at the end of the line ($).
- The > button in the upper right of the dialog will display a popup menu that gives help on Regular Expression syntax.
- Note: The end of the paragraph mark in Regular expression syntax is “\r”, so you can have the paragraph marker (i.e. ¶) included in your search by using “\r” in your “Find what” string. Note, however, that certain documents (e.g. Shoebox/Toolbox files) terminate the end of a line of text with both the carriage return (i.e. “\r”) and a new line (i.e. “\n”). So if you try to do a search to include the paragraph end marker in such a document, you need to include both escape codes in your search string (e.g. “\r\n”). Also note that when the text you want to search for contains a back-slash character (ie. “\”), you have to prefix it with another back-slash to distinguish it from a regular expression meta-character. See the help file on Regular Expressions for details (i.e. Start button, **All Programs / SIL Converters / Help / ICU Regular Expression Help**).

> This utility is also useful when searching for certain dependent Unicode characters which the normal Word search dialog does not support. For example, in the Devanagari range of Unicode, certain vowel characters are considered dependent on a preceding consonant. The regular search dialog in Word does not allow you to search for an occurrence of the dependent vowel character without a preceding independent consonant. With the Regular Expression search utility, however, it is possible to search for any Unicode character regardless of whether it is independent or not.

- **Check round-trip conversion:** this command launches a checking process that goes through the document word-by-word and allows you to check a bi-directional converter from the system repository. This can be useful when developing bi-directional TECkit maps for encoding conversion and/or transliteration.
- The checking is done via a round-trip dialog box that shows the result of the forward conversion as well as the result of the reverse conversion:


SIL Converters for Office: 'Annapurna' using 'Annapurna<>UNICODE'

Input: इक्की 067 077 076 234 032

Forward: इक्‍की 0907 0915 094D 200D 0915 0940 0020

Round-trip: इक्की 067 077 076 235 032

This box shows the result of the round-trip conversion

☑ Skip Identical Forms

Figure 18: Round-trip dialog in Microsoft Word

The checking process will go through the document word-by-word and for each unique font encountered, the program will ask if you want to check the data in that font. If you click Yes , then the Select Converter dialog box will be displayed so you can choose the converter to use. Then each word will be converted in the forward and then reverse direction and the results will be displayed in the Round-trip dialog for you to verify the results.

You can check the **Skip Identical Forms** box in order to skip over any words for which the Input and Round-trip values are the same, but beware that just because the values are the same doesn't mean that the conversion was correct.

See the section above on Microsoft Publisher for a description of the different buttons on this dialog box.

- **Reset:** this command will reset any of the previously issued commands. You can use this, for example, to restart the process with a different converter.

Loading the Add-in in Word 2003

If the SILConverters for Office add-in is not loaded when Microsoft Word 2003 starts (for example, if you are in a different Windows user account than when the SILConverters for Office feature was first installed), follow these steps:

1. Click Tools / Customize...

   The **Customize** dialog will be displayed.

2. In the **Categories** list, select **Tools**, and in the **Commands** list, click and drag the **COM Add-Ins...** item to your **Tools** menu.

3. Close the **Customize** dialog.

4. Click **Tools**, **COM Add-Ins...**

   The COM Add-Ins dialog will be displayed

5. Check the box next to the **SILConverters for Office** item and click OK .

If the **SILConverters for Office** item is not in the COM Add-Ins list, then click the **Add** button to browse for and load the file: C:\Program Files\SIL\SILConverters\SILConvertersOfficeShim.dll.

If the command is successful, you will see the following menu added to the **Word Menu** area:

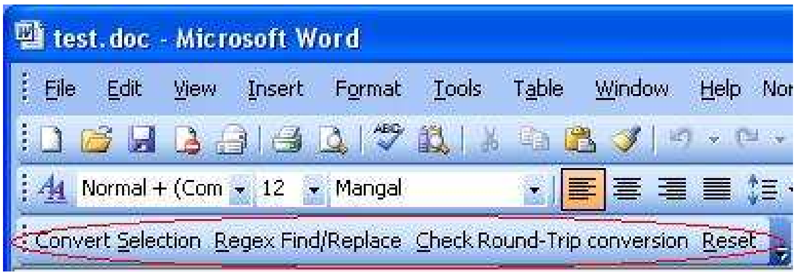


Figure 19: SILConverters Menu in Microsoft Word 2003

If this menu is not visible, it might also means that your security settings may be preventing the add-in from loading. To enable the add-in, see the instructions above in the section on Microsoft Publisher.

If the checkbox next to the "SILConverters for Office" add-in in the COM Add-ins dialog is checked and the Load Behavior message below says "Load on Demand, (not currently loaded)", you can load it by first unloading it and then loading it again. If you want to have the add-in load when Word initially starts, change the value of the following registry key to **3**: HKEY_CURRENT_USER\Software\Microsoft\Office\Word\Addins\SILConvertersOffice.Connect [LoadBehavior]

If this doesn't solve the problem, insure that the following updates are installed on your machine: Control Panel, Add/Remove Programs, **Shared Add-in Extensibility Update for Microsoft .NET Framework 2.0 (KB908002)** and **Shared Add-in Support Update for Microsoft .NET Framework 2.0 (KB908002)**. If these are not presently installed (which can happen if you don't use the master installer to install SILConverters), then run the following program to install them: C:\Program Files\SIL\SILConverters\MS KB908002 Fix\setup.exe.

Loading the Add-in in Word 2007

If the SILConverters for Office add-in is not loaded when Microsoft Word 2007 starts (for example, if you are in a different Windows user account than when the SILConverters for Office feature was first installed), follow these steps:

1. Click the *Office Button* at the top left of the Word 2007 interface (i.e.).
2. Click the **Word Options** button at the bottom of the popup menu (i.e. Word Options).

   The **Word Options** dialog will be displayed.
3. In the Word Options Dialog box, choose the **Add-Ins** item in the left-hand pane.
4. In the **Manage** area, choose **COM Add-ins**, and then click Go.

The COM Add-Ins dialog will be displayed

5. Check the box next to the **SILConverters for Office** item and click OK.

> If the **SILConverters for Office** item is not in the COM Add-Ins list, then click the **Add** button to browse for and load the file: C:\Program Files\SIL\SILConverters\SILConvertersOfficeShim.dll.

In the Toolbar Ribbon, the **Add-Ins** Ribbon will be added.

6. Click on the **Add-Ins** Ribbon tab.

The SILConverters Custom Menu will be display in the **Add-Ins** Ribbon:

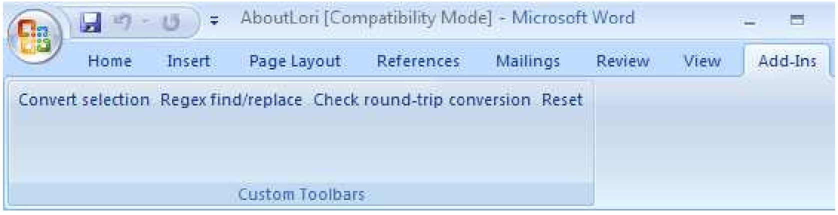


Figure 20b: SILConverters Menu in Microsoft Word 2007

If this menu is not visible, it might also means that your security settings may be preventing the add-in from loading. To enable the add-in, see the instructions above in the section on Microsoft Publisher.

To use in Microsoft Excel:

The SILConverters for Office add-in can be used to convert data in an Excel spreadsheet using converters from the system repository. The add-in is configured to load automatically when Excel starts.

If it loads successfully, you will see the following menu added to the Excel Menu area:

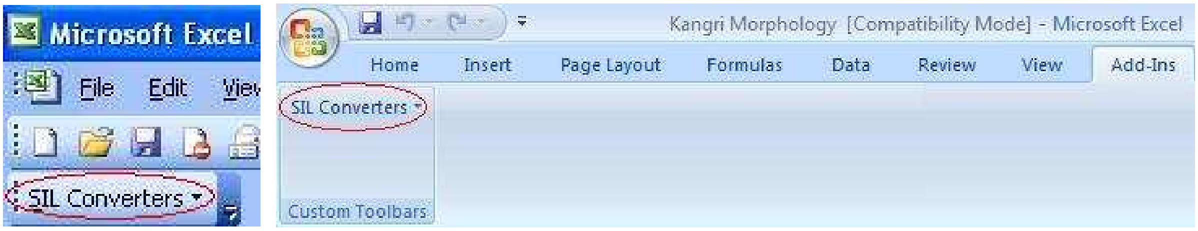


Figure 21: SILConverters menu in Excel 2003 and Excel 2007

> If this menu is not visible, it might also means that your security settings may be preventing the add-in from loading. To enable the add-in, see the instructions above in the section on Microsoft Publisher.

The SILConverters menu in Microsoft Excel contains two menu commands:

- **Convert selection:** this menu command can be used to convert the data in selected cells using a converter from the system repository. To use this command, follow these steps:

A. Select the cells with the data you want to convert

B. Click the **Convert selection** menu command.

The **Select Converter** dialog will be displayed from which you can choose the converter from the system repository to use for converting the selected cells.

> The cells you select can even be from distinct areas of the spreadsheet, but must be on the same spreadsheet for each invocation of this command.

C. Next the add-in will display the **Conversion result** dialog to allow you to confirm the replacement. See the section above on Microsoft Publisher for details of that dialog and the function of the different buttons.

- **Reset:** this command will reset the state of any previously issued Convert selection command. You can use this, for example, to restart the process with a different converter.

There is another way you can use SILConverters in Microsoft Excel: as a custom add-in function.

To enable this functionality in Excel 2003, follow these steps:

1. Click Tools / Add-Ins...

   The **Add-Ins** dialog will be displayed.

2. Click the **Automation button.**

   The **Automation Servers** dialog will be displayed.

3. Select the **SILConvertersConvertersOffice.ConvertFunctions item and click** OK **.**

   You will see a dialog box saying **Cannot find add-in ‘mscoree.dll. Delete from list?** In response to this message box, click NO and then click OK to return to the main Excel window.

To enable this functionality in Excel 2007, follow these steps:

4. Click the *Office Button* at the top left of the Excel 2007 interface (i.e.).
5. Click Excel Options at the bottom of the popup menu.

   The **Excel Options** dialog will be displayed.

6. In the Excel Options Dialog box, choose the **Add-Ins** item in the left-hand pane.
7. In the **Manage** area, choose **Excel Add-ins**, and then click the Go.

   The **Add-Ins** dialog will be displayed.

8. Click the **Automation button.**

   The **Automation Servers** dialog will be displayed.

9. Select the **SILConvertersConvertersOffice.ConvertFunctions item and click** OK **.**

You will see a dialog box saying **Cannot find add-in 'mscoree.dll. Delete from list?** In response to this message box, click NO and then click OK to return to the main Excel window.

To use a converter from the system repository in a cell, follow these steps:

1. Select the cell where you want the converter to put the result of the conversion.
2. Click the **Insert / Function menu.**

   The **Insert Function** dialog will be displayed.
3. In the drop-down box next to where it says, **Or select a category, choose** the **SILConvertersConvertersOffice.ConvertFunctions item**

   The function list window below will list the function ConvertString.
4. Click OK to accept the **ConvertString** function.

   The **Function Arguments** dialog box will be displayed.

   - The **Input** argument is for the cell which contains the string to convert (e.g. A1)
   - The **ConverterName** argument is for the display name of the converter to use (e.g. "SIL IPA93<>UNICODE")
   - The **Forward** argument is a boolean indicating the direction of the conversion (for bi-directional converters). Values are "TRUE" (default) and "FALSE" (for reverse conversion of bi-directional converter).
5. Click OK to accept the **ConvertString** function (e.g. =ConvertString(A1,"Annapurna<>UNICODE")).

   The converted result will be displayed in the cell.

Finally, if after the conversion you want to display the data in another font, you can directly use the Excel Font drop-down box to format the cells with a particular font while they are all still selected.

To use in Microsoft Access:

The SILConverters for Office add-in can be used to convert data in a Microsoft Access database—for example, in a Vocabulary Manager database—using converters from the system repository.

> On some systems the *SILConverters* menu does not appear if you launch Microsoft Access first and then open the database by using the **File / Open** command. If that happens, you can launch Microsoft Access by double-clicking on the Access database file (i.e. *.mdb) in *Windows Explorer*.

Depending on your security settings, the SILConverters for Office add-in may not automatically load when Access starts. To load the add-in in Access 2003, follow these steps:

1. First open the database containing the data you want to convert.
2. Click **Tools / Customize...**

3. The **Customize** dialog will be displayed.
4. In the **Categories** list, select **Tools**, and in the **Commands** list, click and drag the **COM Add-Ins...** item to your **Tools** menu.
5. Close the **Customize** dialog.
6. Click **Tools / COM Add-Ins...**
7. The **COM Add-Ins** dialog will be displayed.
8. Check the box next to the **SILConverters for Office** item and click OK.

   If the checkbox next to the "SILConverters for Office" add-in in the COM Add-ins dialog is checked and the **Load Behavior** message below says "Load on Demand, (not currently loaded)", you can load it by first unloading it and then loading it again.

> If the **SILConverters for Office** item is not in the COM Add-Ins list, then click the **Add** button to browse for and load the file: C:\Program Files\SIL\SILConverters\SILConvertersOfficeShim.dll.

If the command is successful, you will see the following menu added to the Access Menu area:

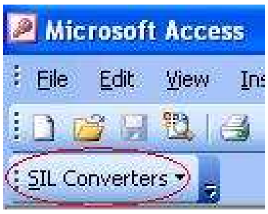


Figure 22: SILConverters menu in Access

If this menu is not visible or if you get an error trying to open the Access database that says, "Microsoft Office Access cannot open due to security restrictions", it means that your security settings are preventing add-in from loading. To enable the add-in, do the following:

- Click **Tools / Macro / Security...**
- The Security dialog will be displayed. To enable the add-in, you can change your Security Level setting to Medium and restart Microsoft Access. When it restarts, the **Security Warning** dialog will be displayed from which you can request to **Enable Macros**.

> If the menu still doesn't show, insure that the following updates are installed on your machine: Control Panel, Add/Remove Programs, **Shared Add-in Extensibility Update for Microsoft .NET Framework 2.0 (KB908002)** and **Shared Add-in Support Update for Microsoft .NET Framework 2.0 (KB908002)**. If these are not presently installed (which can happen if you don't use the master installer to install SILConverters), then run the following program to install them: C:\Program Files\SIL\SILConverters\MS KB908002 Fix\setup.exe.

In Microsoft Access 2007, it is recommended that you double-click on the Access database file (i.e. *.mdb) to launch Access, rather than opening Access first and then opening the file using the **Office Button / Open** command. Also, when the database opens in Access, there may be a *Security Warning* notice as shown in this image:

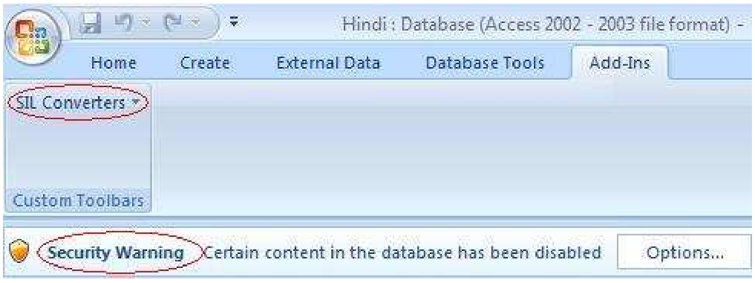

Hindi : Database (Access 2002 - 2003 file format) -
Home
Create
External Data
Database Tools
Add-Ins
SIL Converters
Custom Toolbars
Security Warning
Certain content in the database has been disabled
Options...

Figure 23b: SILConverters menu in Access 2007

Notice the **Add-Ins** tab on which the SIL Converters menu is located.

> You should not dismiss the *Security Warning* notification by clicking the **Options…** button. For some reason, enabling the content causes the *Add-Ins* ribbon to disappear, which makes the SIL Converters menu inaccessible. If this happens, re-launch Access by double-clicking on the database file and ignore the Security Warning. The SILConverters menu commands will still work even if you don’t enable the content.

The SILConverters menu in Microsoft Access contains a single menu command:

- **Convert a field in a table:** this menu command brings up a dialog box listing all of the tables in the opened database. Each table entry contains sub-entries for each of the fields in the table.

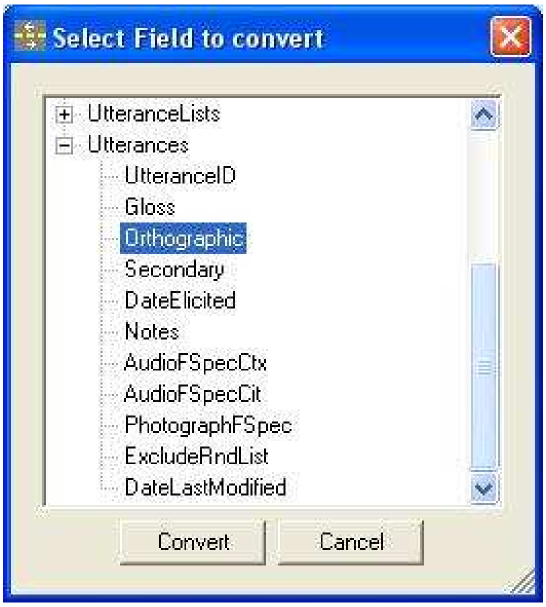


Figure 24: Access Field Select dialog

1. To convert the data in a particular field, just select the field and click the Convert button.

   The **Select Converter** dialog will be displayed from which you can choose the converter to use for the selected field of the database. Then the add-in will display the Conversion result dialog to allow you to confirm the replacement. See the section above on Microsoft Publisher for details of that dialog and the function of the different buttons.

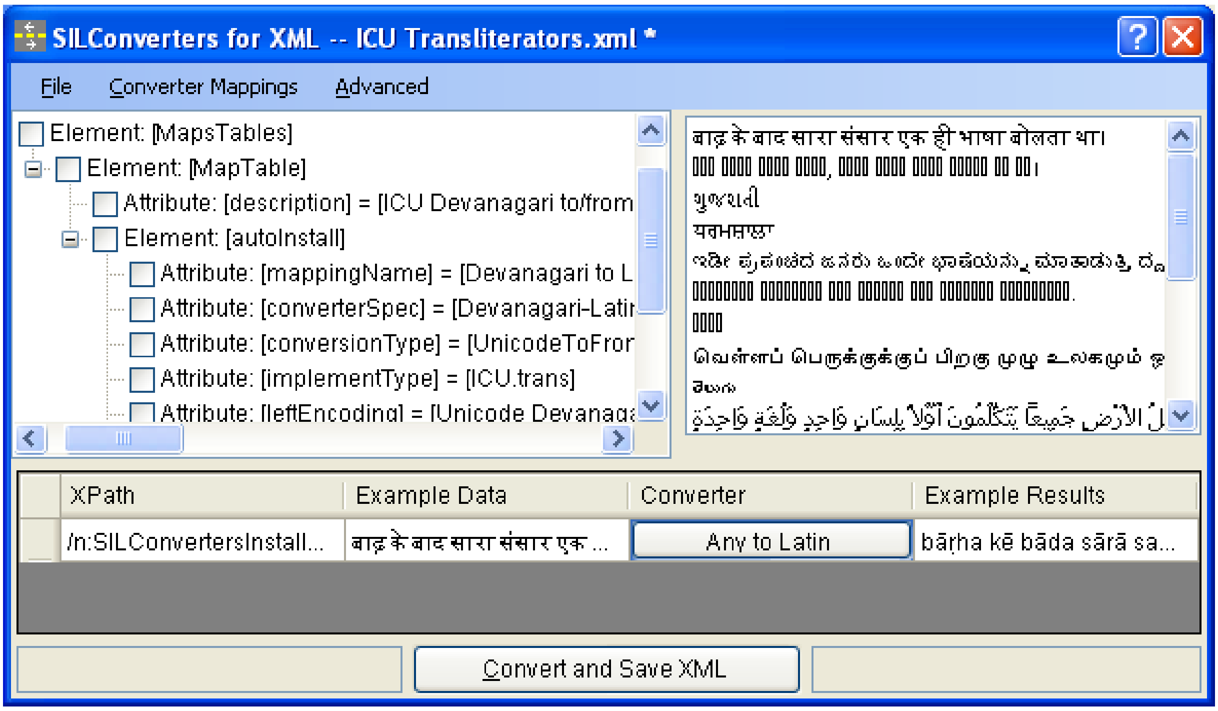
> Note that since this process will convert the data in the database itself, it is important that you back up the database before attempting to convert the data.

XML Data Converter

Use this application to convert the data (attributes or elements) in an XML document using converters from the system repository. This can be used, for example, to convert the data in an Adapt It Knowledge Base from a legacy encoding to Unicode.

To use

- Click **Start / All Programs / SIL Converters / XML Data Converters**.
- For context sensitive help, click ? (top right of program window) and then click the different sections of the main window.



□ Element: [MapsTables]
□ Element: [MapTable]
□ Attribute: [description] = [ICU Devanagari to/from
□ Element: [autoInstall]
□ Attribute: [mappingName] = [Devanagari to L
□ Attribute: [converterSpec] = [Devanagari-Latir
□ Attribute: [conversionType] = [UnicodeToFror
□ Attribute: [implementType] = [ICU.trans]
□ Attribute: [leftEncoding] = [Unicode Devanaga

बाढ़ के बाद सारा संसार एक ही भाषा बोलता था।
□□□ □□□□ □□□□ □□□□, □□□□ □□□□ □□□□ □□□□□ □□ □□।
ગુજરાતી
ਪਰਮਸ਼ਾਲਾ
ಇಡೀ ಪ್ರಪಂಚದ ಜನರು ಒಂದೇ ಭಾಷೆಯನ್ನು ಮಾತಾಡುತ್ತಿದ್ದ
□□□□□□□□ □□□□□□□□ □□□ □□□□□□ □□□ □□□□□□□ □□□□□□□□□.
□□□□
வெள்ளப் பெருக்குக்குப் பிறகு முழு உலகமும் ஒ
తెలుగు
لُ الأَرْضِ جَمِيعاً يَتَكَلَّمُونَ أَوَّلاً بِلِسَانٍ وَاحِدٍ وَلُغَةٍ وَاحِدَةٍ

XPath	Example Data	Converter	Example Results
/n:SILConvertersInstall...	बाढ़ के बाद सारा संसार एक ...	Any to Latin	bāṛha kē bāda sārā sa...

Figure 25: XML Data Converter

The upper left of the program window shows the structure of the XML document (which is very much like an XSD representation of it). In this pane, you can:

- Right-click on a tree item (not the checkbox) and all of the data for that field will be displayed in the list box to the right. You can display either *element* or *attribute* items. However, element items must not have any sub-elements (or there's no data to display).
- Left-click on a tree item (not the checkbox) in order to select that item for conversion using a converter from the system repository.

  The item will appear in the table at the bottom for you to choose the converter to use and view example data from the file.
- If you want to constrain either of the above actions to apply only to certain records in the file, then before clicking on the item, first check the box(es) next to the constraining items. Then when you click on the item to be constrained, you will be presented with a dialog box in which you can specify the constraint value(s).

Note that you can include the same field for conversion more than once with different constraints.

To convert data:

1. Click on the element or attribute item in the upper left window that contains the data that you want to convert (optionally having checked the constraining elements or attributes as discussed above)

   The XPath expression corresponding to the requested item will be displayed in a row of the table at the bottom of the window.

   Data from the file corresponding to the XPath expression will also be displayed in the *Example Data* and the *Example Results* columns.

   You can right-click on a cell in the XPath column to edit the XPath statement.
2. Click on the button in the **Converter** column.

   The *Choose Converter* dialog will be display from which you can select the converter to use for the data.

   Once you click OK, the *Example Results* column will be updated with the results of the conversion using the selected converter.

   If you click on the cells in the *Example Data* column, the program will cycle through the data that matches the XPath expression.
3. Repeat the steps above for all the attributes and elements you want to convert.
4. Click the File / Convert and Save XML Document menu item.

   The **Save As** dialog will be displayed so you can save the converted file.

> It is highly recommended that you do not overwrite the original document unless you are sure that you have an adequate backup of the file.

Example XPath expressions

The following examples show several XPath expressions for different kinds of constraints:

- No filtering:

```
/books/book/title
```

This expression selects the *title* element(s) of the *book* element in the *books* (collection) element

Such an expression would be created by clicking on the *title* element label (not the checkbox) in the tree pane at the top left of the program window.

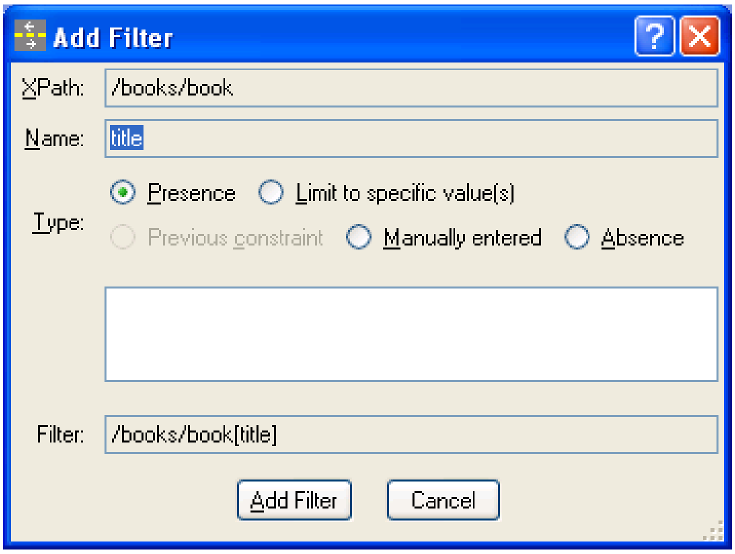
- Element filtering – presence of

```
/books/book[title]
```

selects all *book* elements that contain a *title* element

Such an expression would be created by checking the box next to the *title* element label and then clicking on another label corresponding to the data you're interested in (not the checkbox).

In the resulting dialog, you would choose the **Presence** button as in this image:



Add Filter

XPath: /books/book

Name: title

Type: ● Presence ○ Limit to specific value(s) ○ Previous constraint ○ Manually entered ○ Absence

Filter: /books/book[title]

Figure 26: XPath Filter builder – *Presence of* filter

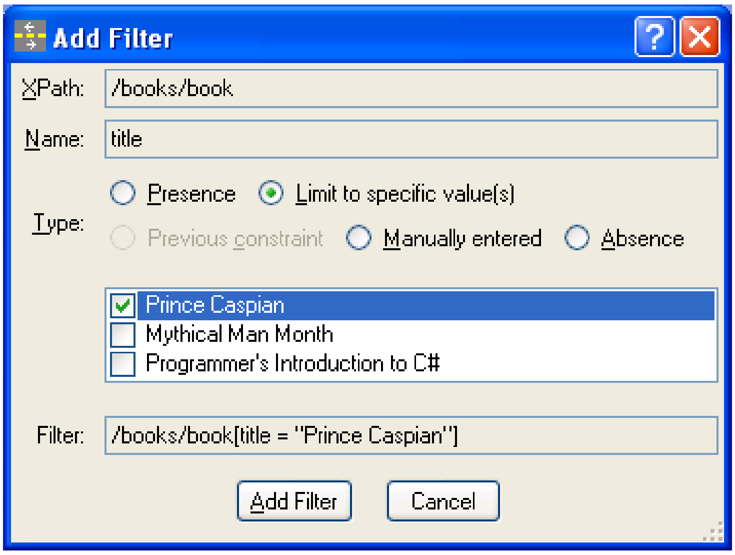
- Element filtering – specific value

```
/books/book[title = 'Prince Caspian']
```

selects all *book* elements that contains a *title* element with the value *'Prince Caspian'*

Such an expression would be created by checking the box next to the *title* element label and then clicking on another label corresponding to the data you're interested in (not the checkbox).

In the resulting dialog, you would choose the **Limit to specific values** button as in this image and check the box(es) corresponding to the values you want to match.



XPath: /books/book

Name: title

Type: ○ Presence ● Limit to specific value(s) ○ Previous constraint ○ Manually entered ○ Absence

☑ Prince Caspian
□ Mythical Man Month
□ Programmer's Introduction to C#

Filter: /books/book[title = "Prince Caspian"]

Figure 27: XPath Filter builder – *Specific Values* filter

- Attribute filtering:

```
/books/book[@author]
```

selects all the *book* elements that have an overt *author* attribute

```
/books/book[@author = 'C. S. Lewis']
```

selects all the *book* elements that have an *author* attribute with a value of *'C. S. Lewis'*

- Multiple conditions:

```
/font[@name = 'SILDoulos IPA93' or @name = 'SILManuscript IPA93' or @name = 'SILSophia IPA93']
```

select the *font* element(s) which have a *name* attribute of *'SILDoulos IPA93'*, *'SILManuscript IPA93'*, or *'SILSophia IPA93'*.

MS Word converters

You can use SILConverters directly in MS Word. The converters are *macros* contained in three Microsoft Word document templates (DOTs). These macros use the EncConverters repository to accomplish different tasks.

- Data Conversion Macro in Data Conversion Macro xxxx.dot
- SpellingFixer in SpellFixer.dot
- Consistency Spelling Checker in Consistent Spelling Checker xxxsc.dot

If you select the WordDOTs feature node in the SILConverters 3.1 installer, the installer will put these templates into your Templates folder (normally C:\Documents and Settings\<user>\Application Data\Microsoft\Templates).

To use

- To access the document template clients from within Microsoft Word, click **Tools… / Templates and Add-Ins**.

The **Templates and Add-Ins** dialog box will be displayed.

- In the Templates and Add-Ins dialog box, click Add.

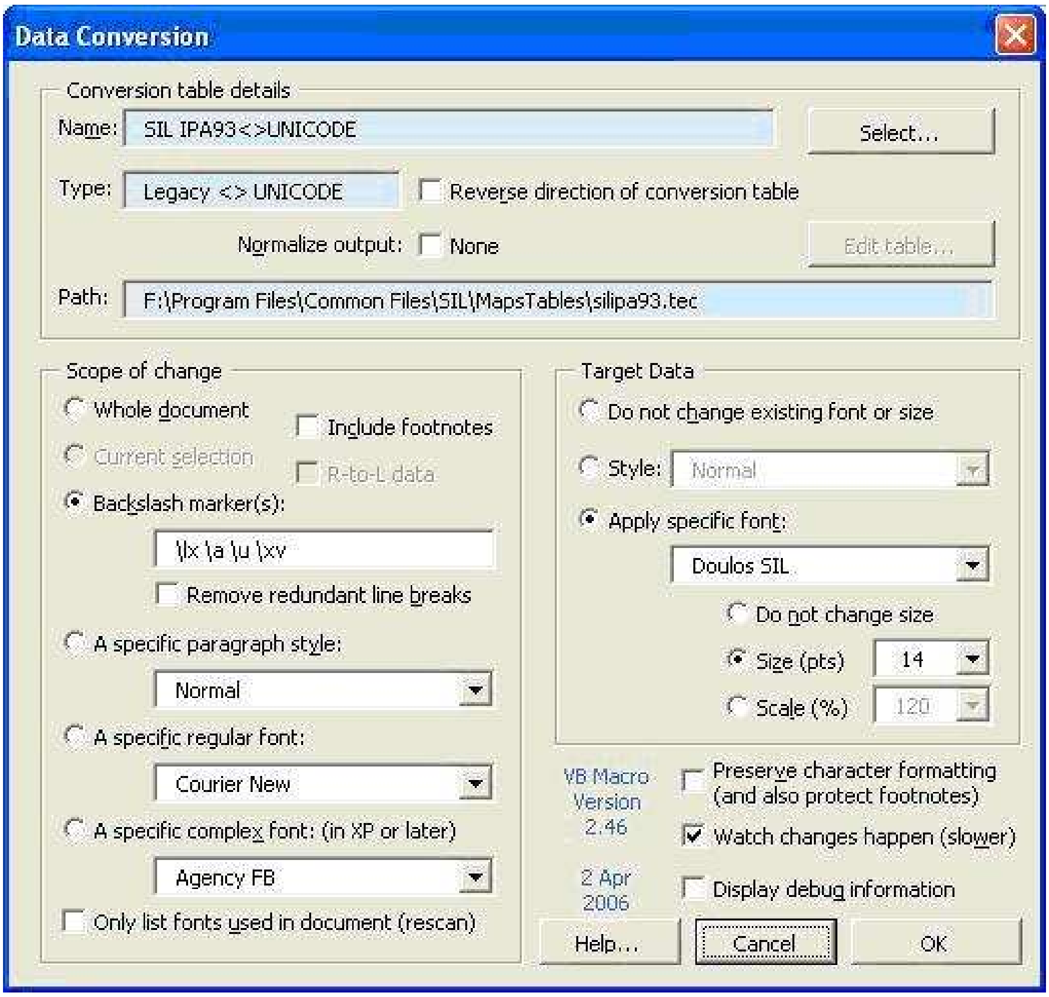
  Your Templates folder will be opened.

- Select the Word .dot file you want to use.

> If multiple users on the machine want to use the document template, you need to manually move the .DOT files to some common location and each user will need to browse for them individually in **Tools / Templates and Add-Ins**. If you want one or more of these document templates to start up automatically when Word starts, move them either to the current user's Startup folder (i.e. C:\Documents and Settings\<user>\Application Data\Microsoft\Word\STARTUP). For all users, put it in the global startup folder (e.g. C:\Program Files\Microsoft Office\OFFICE11\STARTUP).

Data Conversion Macro

Use the Data Conversion Macro to convert text in any arbitrary Word document based on Font name, Style, or even the current selection using converters from the system repository. It also supports SFM documents. Open the document template for more instructions (i.e. Start button, **All Programs / SIL Converters / Help / Help for the Data Conversion Macro**).



Data Conversion

Conversion table details

Name: SIL IPA93<>UNICODE

Type: Legacy <> UNICODE

□ Reverse direction of conversion table

Normalize output: □ None

Path: F:\Program Files\Common Files\SIL\MapsTables\silipa93.tec

Scope of change

○ Whole document

□ Include footnotes

○ Current selection

□ R-to-L data

● Backslash marker(s): \lx \a \u \xv

□ Remove redundant line breaks

○ A specific paragraph style: Normal

○ A specific regular font: Courier New

○ A specific complex font: (in XP or later) Agency FB

□ Only list fonts used in document (rescan)

Target Data

○ Do not change existing font or size

○ Style: Normal

● Apply specific font: Doulos SIL

○ Do not change size

● Size (pts) 14

○ Scale (%) 120

VB Macro Version 2.46

2 Apr 2006

□ Preserve character formatting (and also protect footnotes)

☑ Watch changes happen (slower)

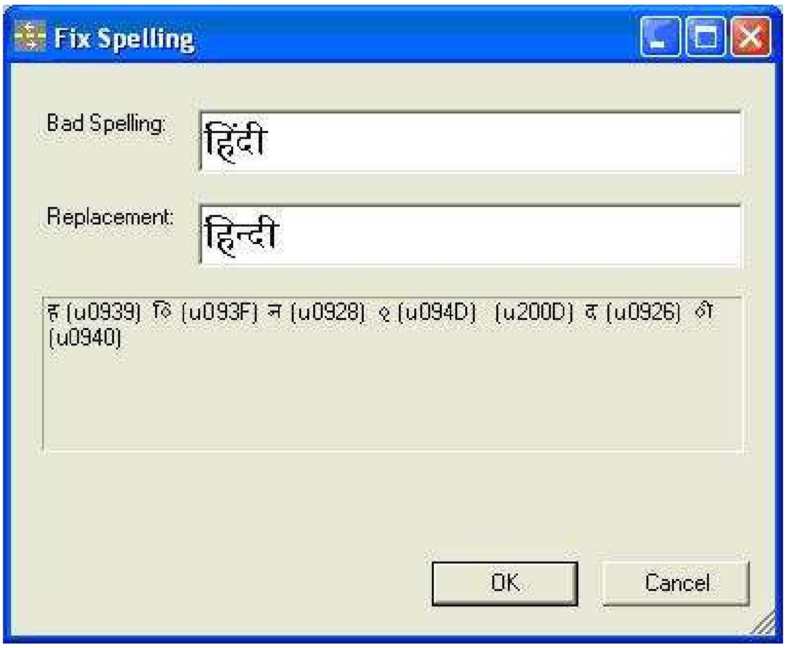
□ Display debug information

Figure 28: Data Conversion Macro dialog

SpellingFixer

Use the SpellingFixer document template to correct misspelled words or make certain orthographic changes based on a user-defined database of bad-good spelling pairs. This is particular useful when you want to:

- Condition spelling changes to be at either a beginning or ending word boundary, or
- Convert a portion of a word (as opposed to full word forms).



Bad Spelling: हिंदी

Replacement: हिन्दी

ह (u0939) ि (u093F) न (u0928) ् (u094D) (u200D) द (u0926) ी (u0940)

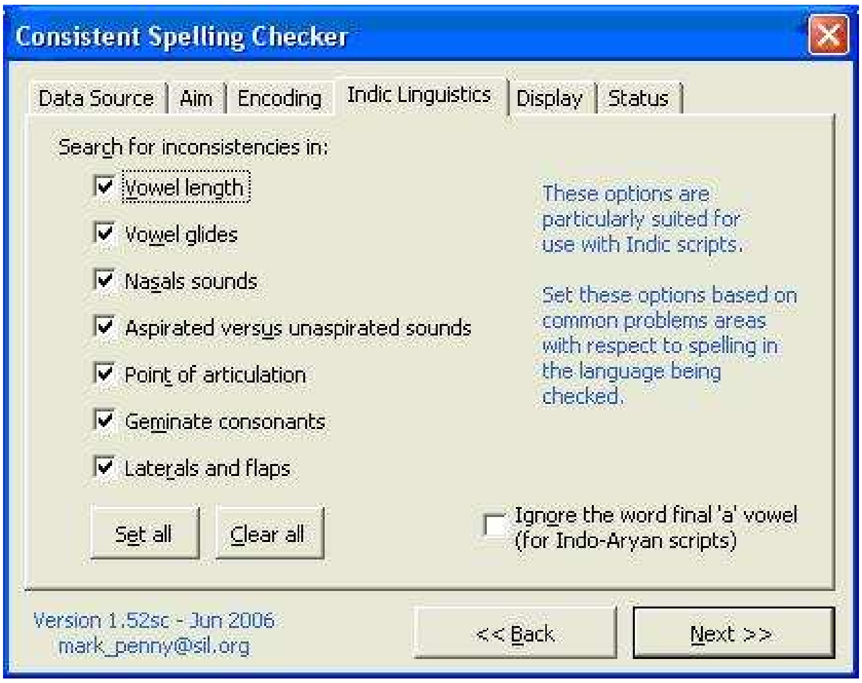
Figure 29: Enter correction rules dialog

Once you have a database of such spelling *fixes* (or *consistent changes*), use one of the **Correct Whole Document** menu commands to go through all the words in the document to search for misspelled words. See the document template for instructions (i.e. Start button, **All Programs / SIL Converters / Help / Help for the SpellFixer macro**).

Consistency Spelling Checker

Use the Consistency Spelling Checker document template for a simple way of working with data (in *any* language, and *any* script) in *Microsoft Word* documents, *Plain Text* files or any *Toolbox* database to:

- Check consistency of spelling (semi-automatically) based on linguistic principles
- Apply global spelling changes:
- to multiple documents which are currently open
- by generating a CC table of changes to be applied to one or more plain text databases (such as Toolbox files)
- Create a character inventory with frequency count
- Create unique wordlists from one or more Word documents as:
- a *Word* document table with frequency counts, or
- a Toolbox (MDF-formatted) database for starting a lexicon



Search for inconsistencies in:

☑ Vowel length

☑ Vowel glides

☑ Nasals sounds

☑ Aspirated versus unaspirated sounds

☑ Point of articulation

☑ Geminate consonants

☑ Laterals and flaps

These options are particularly suited for use with Indic scripts.

Set these options based on common problems areas with respect to spelling in the language being checked.

□ Ignore the word final 'a' vowel (for Indo-Aryan scripts)

Version 1.52sc - Jun 2006
mark_penny@sil.org

Figure 30: Spelling inconsistency parameter dialog

> This tool is *not* a full-fledged spelling checking tool. It does not use language-specific dictionaries, and therefore knows nothing about the languages it checks. It is only a *consistency checking tool* based on phonological similarity, or sets of user-defined ambiguous characters.

Prerequisites

The Spelling Consistency Checker macro requires that you install this software:

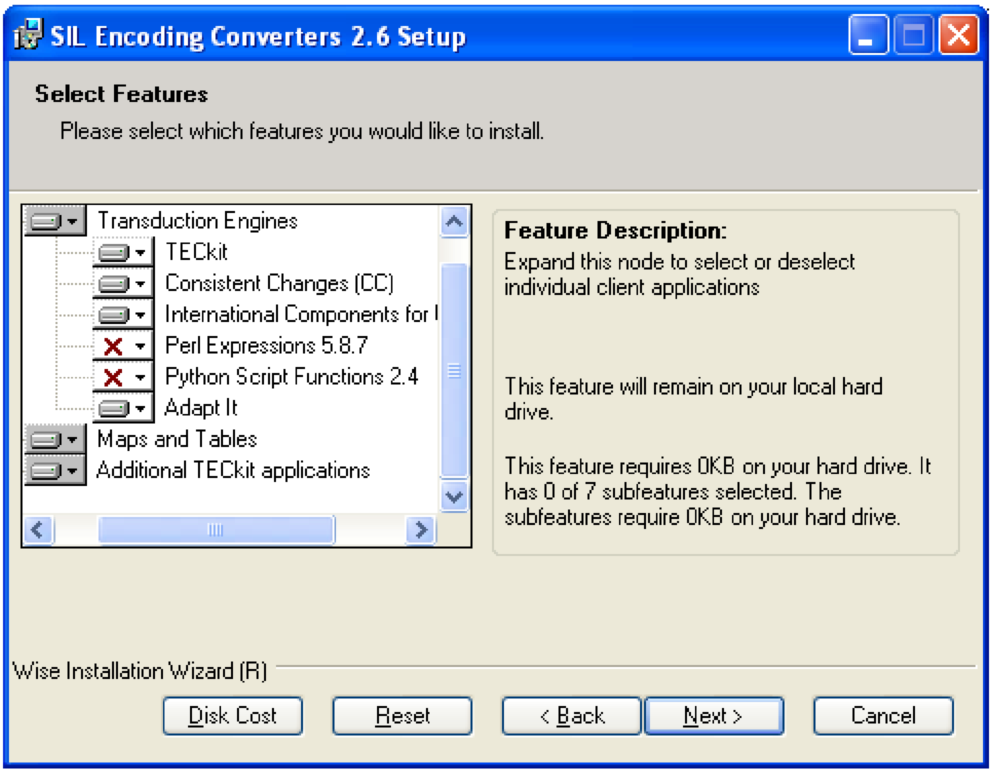
- Operating system: Windows XP (or newer)
- Word XP (or 2003)
- The Field Linguist's Toolbox http://www.sil.org/computing/toolbox/, (formerly called Shoebox) to work from a Word-list
- Consistent Changes for Windows http://www.sil.org/computing/catalog/cc.html (CCW32.exe) to apply global changes

See the document template for instructions (i.e. Start button, **All Programs / SIL Converters / Help / Help for the Consistent Spelling Checker**).

SIL Converters’ Transduction Engines

Several of the transduction engines in Figure 1 are provided by the EncConverters’ repository object itself (i.e. the code page converter and the Compound and Primary-Fallback meta converters) and are always available. The rest depend on external programs (SIL and other Open Source programs) and installation is optional, depending on your need.

Most end users will not need to concern themselves with these details except to be sure that the necessary transduction engine is installed for the converters they want to use. Chances are that someone in your entity has already created a map file that you can use to convert the encoding of your data. In this situation, you need to be sure that you install the proper transduction engine required by the map or table that implements the conversion you want.



SIL Encoding Converters 2.6 Setup

Select Features

Please select which features you would like to install.

- Transduction Engines
  - TECkit
  - Consistent Changes (CC)
  - International Components for I
  - ✗ Perl Expressions 5.8.7
  - ✗ Python Script Functions 2.4
  - Adapt It
- Maps and Tables
- Additional TECkit applications

Feature Description:

Expand this node to select or deselect individual client applications

This feature will remain on your local hard drive.

This feature requires 0KB on your hard drive. It has 0 of 7 subfeatures selected. The subfeatures require 0KB on your hard drive.

Wise Installation Wizard (R)

Figure 31: Optional SIL Converters' Transduction Engines

TECkit

Other applications use TECkit, a low-level toolkit, to perform encoding conversions (e.g., when importing legacy data into a Unicode-based application). The primary component of the TECkit package is a library that performs conversions. This is the "TECkit engine". The engine relies on mapping tables in a specific binary format (see TECkit documentation). A compiler creates such tables from a human-readable mapping description (a simple text file).

In EncConverters, you can select either the compiled *.tec file or the uncompiled, human-readable *.map) to be the converter. If you choose the latter, EncConverters will automatically compile an out-of-date .tec file when it is used to convert data.

See *Adding Converters to the System Repository* below for details about adding TECkit maps to the system repository.

Consistent Changes (CC)

Use Consistent Change tables to find all occurrences of specified characters, words, or phrases in a string of text, and then change them in a consistent way. The change may be done in every occurrence or only when certain conditions are met.

CC is like the find-and-replace feature in a text editor, but much more powerful. It allows you

- to make changes which take context into consideration, and
- to make a whole set of changes at once.

SpellFixer is also available. This is a user-friendly graphical user interface for creating consistent change tables. This interface is primarily available via the SpellFixer.dot Microsoft Word document template mentioned above in SILConverters' client applications.

See *Adding Converters to the System Repository* below for details about adding CC tables to the system repository.

International Components for Unicode (ICU) 4.0

Three distinct EncConverters-related features as well as other features of ICU used by other client are applications that must be installed as a unit.

For SILConverters, three transduction engines are included in this feature:

- ICU Transliterators: provides a series of transliterators for various ranges of Unicode (c.f. Devanagari to Latin) as well as the ability to write custom rules for doing transliteration. See http://icu.sourceforge.net/userguide/Transform.html for more details on the use and syntax of ICU Transliterators.
- ICU Converters: provides comprehensive character set conversion services, mapping tables, and implementations for many encodings. Since ICU uses Unicode (UTF-16) internally, all converters convert between UTF-16 (with the *endianness* according to the current platform) and another encoding. This converter includes other Unicode encodings. These are typically of more interest to programmers than end-users. See http://icu.sourceforge.net/userguide/conversion.html for more details on ICU converters.
- Regular Expression: provides applications with the ability to apply regular expression matching to Unicode string data. The regular expression patterns and behavior are based on Perl's regular expressions. See http://icu.sourceforge.net/userguide/regexp.html for more details on the syntax of ICU Regular Expressions.

Perl Expressions 5.10.1

The Perl Expressions 5.10.1 transduction engine allows you to write Perl expressions to do text processing in EncConverter client applications.

> This feature requires installation of a separate Perl 5.10.1 distribution to be installed.
>
> The Perl plug-in has been tested with the freely available Perl distribution from ActiveState Perl at: http://www.activestate.com/activeperl.

Python Script Functions 2.5

The *Python Script Functions 2.5* transduction engine allows you to do text processing using Python functions in EncConverter client applications.

> This feature requires a separate Python 2.5 distribution to be installed.
>
> The Python plug-in has been tested with the following freely available Python distributions: ActiveState Python at http://www.activestate.com/solutions/python or Python.org at: http://www.python.org/download/.

Adapt It

The Adapt It package contains two transduction engines:

- AdaptIt Knowledge Base Lookup Converter

  This converter allows you to use an AdaptIt Knowledge Base to provide data for a whole-word-form lookup/replacement transduction process. This can be used by client applications that want to access the database of Source to Target word mappings in an existing AdaptIt Knowledge Base.

  See the help file **Start / All Programs / SIL Converters / Help / Help for AdaptIt Plug-in** for more details.

- AdaptIt Target Word Guesser

  This transducer can be used in an AdaptIt adaptation project to guess target words based on the Source/Target word pairs in the AdaptIt Knowledge Base.

  See the help file **Start / All Programs / SIL Converters / Help / Help for AdaptIt Target Word Guesser** for more details.

> You must use AdaptIt v 3.1 or newer and have it configured to save the knowledge base as an .XML file in order for these transducers to be able to read it.
>
> For non-Unicode Adapt It projects, you must use AdaptIt v 3.4.1 or newer.

Additional TECkit applications

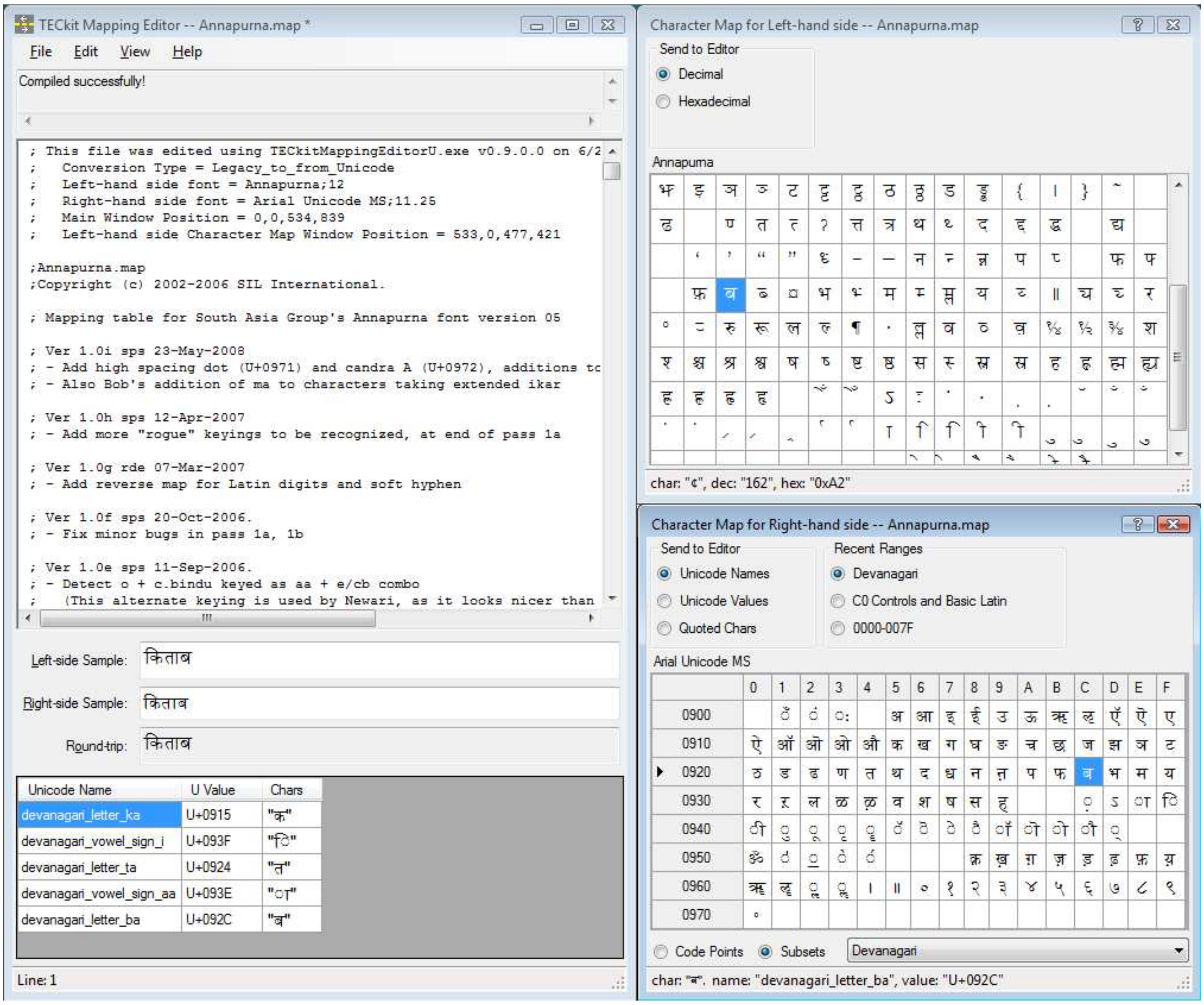
TECkit Map Unicode Editor

The *TECkit map Unicode Editor* is one more SILConverters' client application shown in Figure 1. Use this program to develop TECkit maps for encoding conversion or other text processing applications (e.g. Transliteration).

Steps

1. Install this application by selecting the **TECkit Map Unicode Editor** feature under the Additional TECkit applications feature node.
2. Start the program by clicking **Start… / All Programs / SIL Converters / TECkit / TECkit Map Unicode Editor**.

For context sensitive help, select different sections of the main window and press the F1 key or click ? (top right of program window) and then click on the different sections of the program window.


TECkit Mapping Editor -- Annapurna.map *
File
Edit
View
Help
Character Map for Left-hand side -- Annapurna.map
Send to Editor
Annapurna
Character Map for Right-hand side -- Annapurna.map
Send to Editor
Recent Ranges
Arial Unicode MS

Compiled successfully!

```
; This file was edited using TECkitMappingEditorU.exe v0.9.0.0 on 6/2
;   Conversion Type = Legacy_to_from_Unicode
;   Left-hand side font = Annapurna;12
;   Right-hand side font = Arial Unicode MS;11.25
;   Main Window Position = 0,0,534,839
;   Left-hand side Character Map Window Position = 533,0,477,421

;Annapurna.map
;Copyright (c) 2002-2006 SIL International.

; Mapping table for South Asia Group's Annapurna font version 05

; Ver 1.0i sps 23-May-2008
; - Add high spacing dot (U+0971) and candra A (U+0972), additions tc
; - Also Bob's addition of ma to characters taking extended ikar

; Ver 1.0h sps 12-Apr-2007
; - Add more "rogue" keyings to be recognized, at end of pass 1a

; Ver 1.0g rde 07-Mar-2007
; - Add reverse map for Latin digits and soft hyphen

; Ver 1.0f sps 20-Oct-2006.
; - Fix minor bugs in pass 1a, 1b

; Ver 1.0e sps 11-Sep-2006.
; - Detect o + c.bindu keyed as aa + e/cb combo
;   (This alternate keying is used by Newari, as it looks nicer than
```

Left-side Sample: किताब

Right-side Sample: किताब

Round-trip: किताब

Unicode Name	U Value	Chars
devanagari_letter_ka	U+0915	"क"
devanagari_vowel_sign_i	U+093F	"ि"
devanagari_letter_ta	U+0924	"त"
devanagari_vowel_sign_aa	U+093E	"ा"
devanagari_letter_ba	U+092C	"ब"

Line: 1

● Decimal
○ Hexadecimal

char: "¢", dec: "162", hex: "0xA2"

● Unicode Names
○ Unicode Values
○ Quoted Chars

● Devanagari
○ C0 Controls and Basic Latin
○ 0000-007F

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0900		◌ँ	◌ं	◌ः		अ	आ	इ	ई	उ	ऊ	ऋ	ऌ	ऍ	ऎ	ए
0910	ऐ	ऑ	ऒ	ओ	औ	क	ख	ग	घ	ङ	च	छ	ज	झ	ञ	ट
0920	ठ	ड	ढ	ण	त	थ	द	ध	न	ऩ	प	फ	ब	भ	म	य
0930	र	ऱ	ल	ळ	ऴ	व	श	ष	स	ह			◌़	ऽ	◌ा	◌ि
0940	◌ी	◌ु	◌ू	◌ृ	◌ॄ	◌ॅ	◌ॆ	◌े	◌ै	◌ॉ	◌ॊ	◌ो	◌ौ	◌्		
0950	ॐ	◌॑	◌॒	◌॓	◌॔				क़	ख़	ग़	ज़	ड़	ढ़	फ़	य़
0960	ॠ	ॡ	◌ॢ	◌ॣ	।	॥	०	१	२	३	४	५	६	७	८	९
0970	॰															

○ Code Points ● Subsets Devanagari

char: "ब". name: "devanagari_letter_ba", value: "U+092C"

Figure 32: TECkit Map Unicode Editor

The left portion of the program window is the editor for the TECkit map and the right portion gives some short cut ways of determining the code points of various letters in the selected fonts.

You can

- Type (or paste) text into the Sample boxes in the lower left window.

  The code point values and/or names of the characters in the string will be displayed in the table below those boxes.

- Click the cells in that table to insert those values into the map (in the upper left).

- To save your map to the default system repository, click **File... / Add to System Repository**, navigate to C:\Program Files\Common Files\SIL\MapsTables\, and click Save .

The *TECkit Map Transduction Engine* dialog box will be displayed.

Tips

- The font glyphs for the left and right-hand side fonts will be displayed in the two tables on the right-hand side. Click on a cell in either of those tables to insert their values into the map or Ctrl + Click to insert them into the Sample boxes. For Unicode encodings, you can also limit the table to just a sub-set of the Unicode character values by using the combo box selector at the bottom right of Figure 30 (which currently shows the **Devanagari** range)
- The map is automatically compiled as you make changes and you can click errors in the compiler results window (Figure 32, extreme top left) to jump to problem statements.
- The program will automatically convert the data in the Left-side Sample box (in the forward direction) in order to check conversion as you work on it. It will also convert the Right-side Sample (in the reverse direction) in order to check the round-trip capability of the map.
- Context-sensitive help is available—select a portion of the window and press F1 or click ? (right-hand window).

Adding Converters to the System Repository

There are two primary ways of adding converters to the System Repository, by using either the

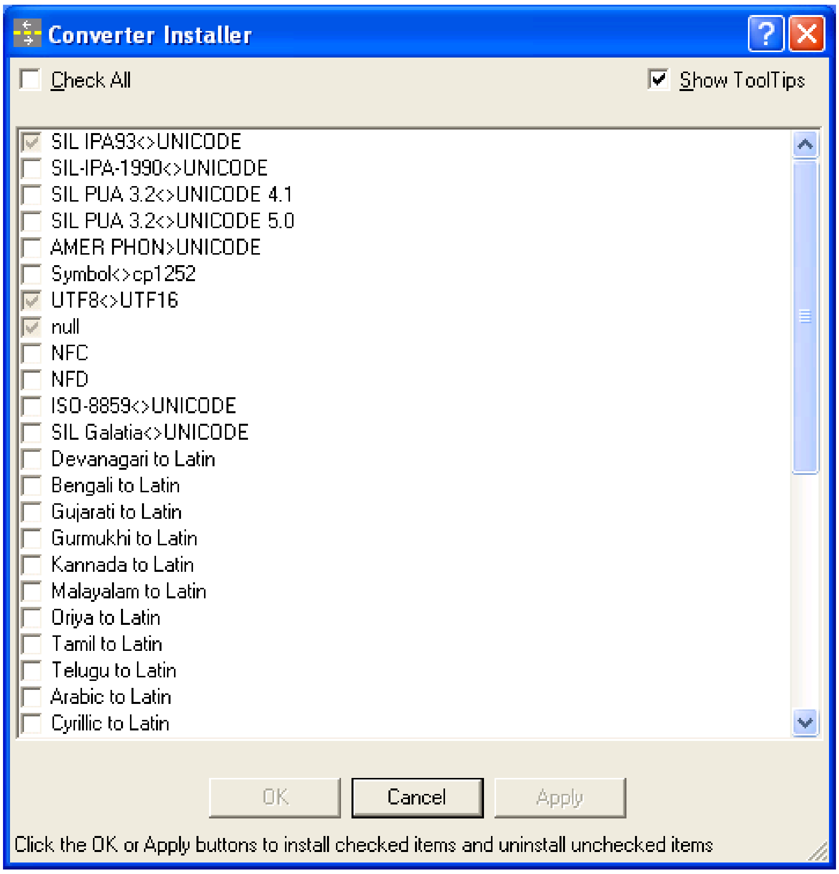
- Converter Installer or
- Transduction Engine configuration dialogs.

Converter Installer

If the converter you want to install into the system repository comes as part of the Maps and Tables features in the SILConverters installer (e.g. the SIL IPA93<>UNICODE converter that comes as part of the Basic Converters package), you can install it into the system repository by running the Converter Installer application.

How to get there

- When running the Master Installer, this utility automatically launches as the last item in the installer sequence.
- From the Windows taskbar, click **Start… / All Programs / SIL Converters / Launch Converter Options Installer**.
- From the Clipboard EncConverter popup by selecting **Launch Converter Installer**.



□ Check All
☑ Show ToolTips

☑ SIL IPA93<>UNICODE
□ SIL-IPA-1990<>UNICODE
□ SIL PUA 3.2<>UNICODE 4.1
□ SIL PUA 3.2<>UNICODE 5.0
□ AMER PHON>UNICODE
□ Symbol<>cp1252
☑ UTF8<>UTF16
☑ null
□ NFC
□ NFD
□ ISO-8859<>UNICODE
□ SIL Galatia<>UNICODE
□ Devanagari to Latin
□ Bengali to Latin
□ Gujarati to Latin
□ Gurmukhi to Latin
□ Kannada to Latin
□ Malayalam to Latin
□ Oriya to Latin
□ Tamil to Latin
□ Telugu to Latin
□ Arabic to Latin
□ Cyrillic to Latin

Click the OK or Apply buttons to install checked items and uninstall unchecked items

Figure 33: Converter Options Installer

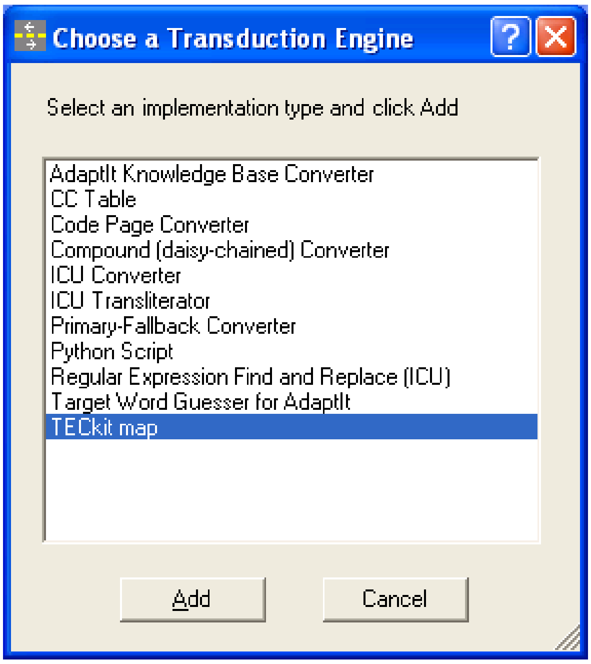
Installing converters

- To install one or more of these converters into the system repository, check the box next to the converters you want and click OK or Apply .
- To remove converters from the system repository, clear the check box and click OK or Apply.

For detailed instructions see the Converter Installer section in the installation documentation.

Choose a Transduction Engine dialog box

If you have your own converter map (e.g. created with the TECkit map Unicode Editor) or one given to you not as part of an installer feature, you can add it to the system repository via the **Choose a Transduction Engine** dialog box.



Select an implementation type and click Add

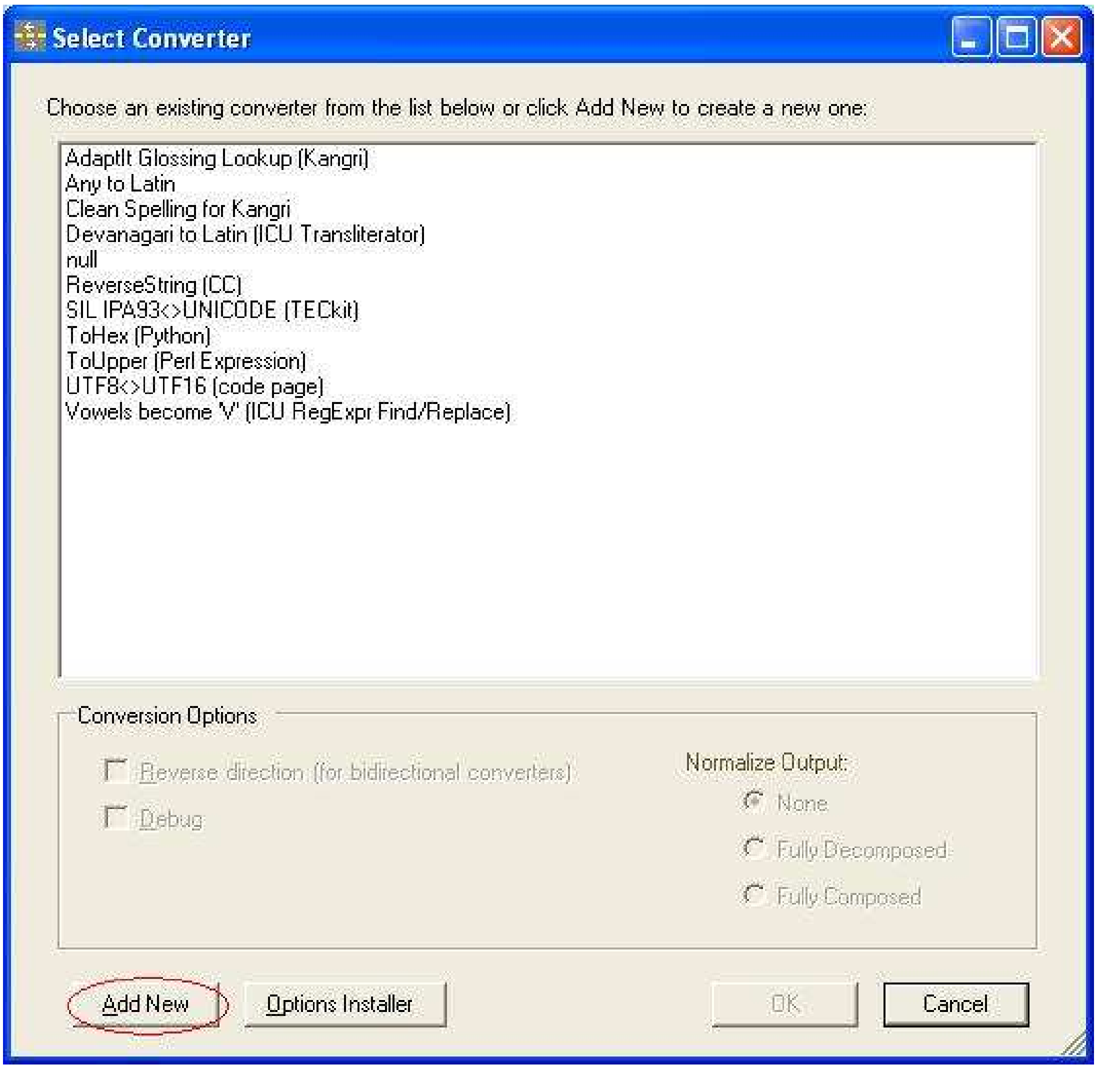
AdaptIt Knowledge Base Converter
CC Table
Code Page Converter
Compound (daisy-chained) Converter
ICU Converter
ICU Transliterator
Primary-Fallback Converter
Python Script
Regular Expression Find and Replace (ICU)
Target Word Guesser for AdaptIt
TECkit map

Figure 34: Choose Transduction Engine dialog

- Select the item from the list that matches the type of converter you want to add and click Add .

How to get there

- If you are running the *Clipboard EncConverter* application, right-click the icon in the system tray and choose the **Add Converter** item.
- If you are using *AdaptIt*, *Data Conversion Macro*, *Bulk SFM Converter*, *XML Data Converter*, or *SpellFixer*; open the **Select Converter** dialog window (Figure 35) and click Add New .



Select Converter

Choose an existing converter from the list below or click Add New to create a new one:

AdaptIt Glossing Lookup (Kangri)
Any to Latin
Clean Spelling for Kangri
Devanagari to Latin (ICU Transliterator)
null
ReverseString (CC)
SIL IPA93<>UNICODE (TECkit)
ToHex (Python)
ToUpper (Perl Expression)
UTF8<>UTF16 (code page)
Vowels become 'V' (ICU RegExpr Find/Replace)

Conversion Options

☐ Reverse direction (for bidirectional converters)

☐ Debug

Normalize Output:

● None

○ Fully Decomposed

○ Fully Composed

Figure 35: Select Converter dialog

Transduction Engine Details

TECkit map

To add a TECkit map to the system repository, select *TECkit map* from the *Choose a Transduction Engine dialog box*, and click Add.

The TECkit map Setup dialog will be displayed.

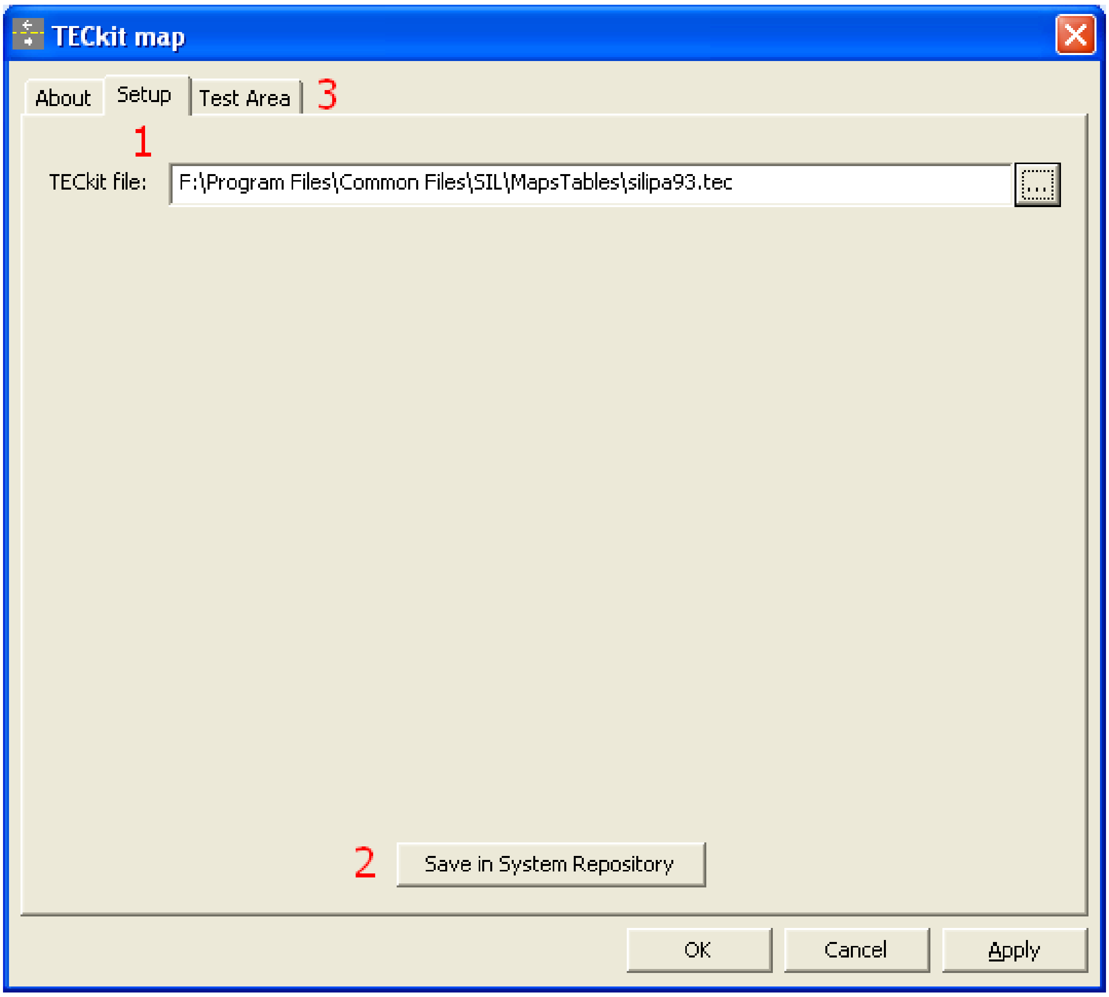

TECkit map
About
Setup
Test Area
3
1
TECkit file:
F:\Program Files\Common Files\SIL\MapsTables\silipa93.tec
...
2
Save in System Repository
OK
Cancel
Apply

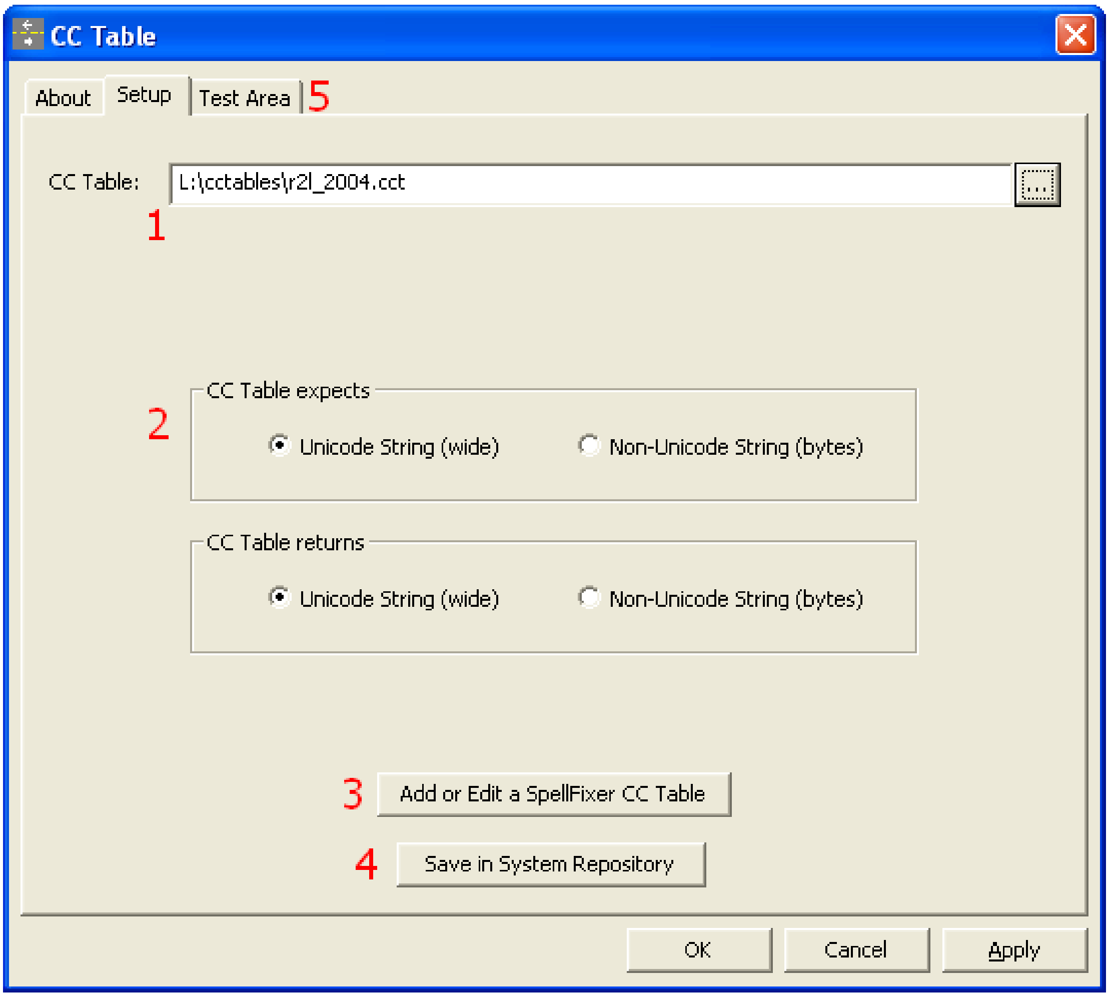
Figure 36: TECkit Setup

1. Browse with the ... button for the TECkit .map or .tec file.
2. To permanently add the converter to the System Repository, click Save in System Repository .
3. Click **Test Area** to test the converter with some sample data.

Consistent Changes (CC)

To add a CC table to the system repository, select **CC table** from the *Choose a Transduction Engine dialog box* and click Add .

The CC table Setup dialog will be displayed:



CC Table: L:\cctables\r2l_2004.cct

CC Table expects
● Unicode String (wide) ○ Non-Unicode String (bytes)

CC Table returns
● Unicode String (wide) ○ Non-Unicode String (bytes)

Figure 37: CC Table Setup

1. Browse with the ... button for the CC table file.
2. For **CC Table expects** and **CC Table returns**, select the desired encoding and click Apply .

Tip:

If it expects Unicode-encoded data, select that option or your data may be incorrectly converted. For Non-Unicode (byte) data, the default system code page will be used to convert your data when necessary.

3. If you installed the SpellFixer plug-in, click **Add or Edit a SpellFixer CC Table**.

   This will allow you to create or edit an existing SpellFixer project.

Tip:

Though primarily a Microsoft Word-based tool, you can use The SpellFixer application to create a CC table. Use the SpellFixer graphical user-interface to configure *Bad Spelling* and *Good Spelling* pairs, which then are put into a CC table. The Microsoft Word document template also has macros for processing the text in a file in a word-by-word manner so you can use it in a Find First/Next fashion to correct spelling errors. The SpellFixer.dot file has further usage information.

4. To permanently add the converter to the System Repository, click Save in System Repository .

You do not need to add a SpellFixer project to the System Repository, since it will be added automatically by the Project editor.

5. Click **Test Area** to test the converter with some sample data.

ICU Transliterator

- To add an ICU Transliterator to the system repository, choose **ICU Transliterator** from the *Choose a Transduction Engine dialog box* and click Add .

  The ICU Transliterator Setup dialog will be displayed:

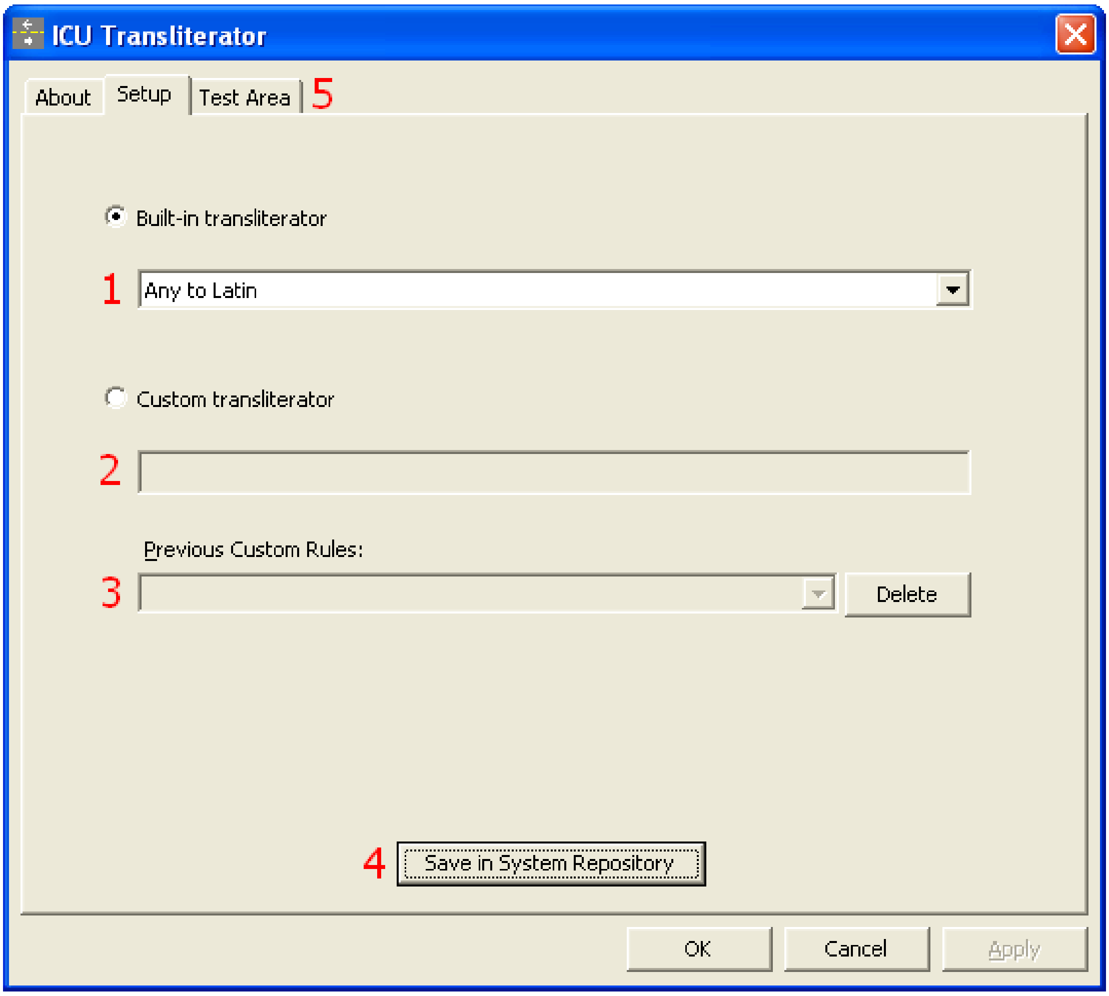


Figure 38: ICU Transliterator Setup

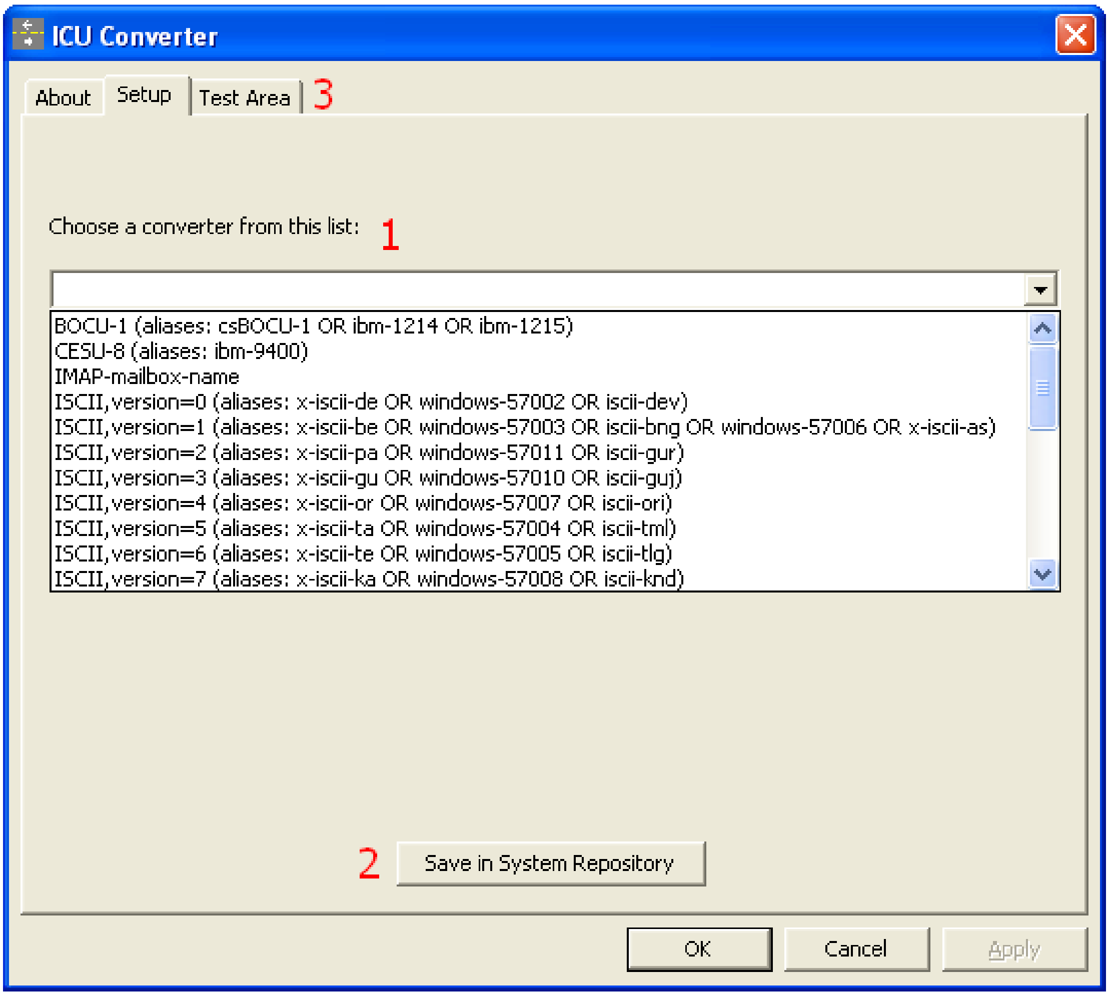
1. To use one of the built-in transliterators, choose **Built-in transliterator**.
2. To write a custom transliterator using the syntax described on the webpage referred to above, choose **Custom transliterator** and enter the transliterator syntax in the box.

   Previous Custom Rules will be enabled, showing examples of useful custom rules and others that you wrote (Figure 38, number 3).
3. Click Delete to remove unwanted rules.
4. To permanently add the converter to the System Repository, click Save in System Repository .
5. Click **Test Area** to test the converter with some sample data.

ICU Converters

- If you want to add an ICU Converter to the system repository, choose **ICU Converter** from the *Choose a Transduction Engine dialog box* and click Add .

The ICU Converter Setup dialog will be displayed:



BOCU-1 (aliases: csBOCU-1 OR ibm-1214 OR ibm-1215)
CESU-8 (aliases: ibm-9400)
IMAP-mailbox-name
ISCII,version=0 (aliases: x-iscii-de OR windows-57002 OR iscii-dev)
ISCII,version=1 (aliases: x-iscii-be OR windows-57003 OR iscii-bng OR windows-57006 OR x-iscii-as)
ISCII,version=2 (aliases: x-iscii-pa OR windows-57011 OR iscii-gur)
ISCII,version=3 (aliases: x-iscii-gu OR windows-57010 OR iscii-guj)
ISCII,version=4 (aliases: x-iscii-or OR windows-57007 OR iscii-ori)
ISCII,version=5 (aliases: x-iscii-ta OR windows-57004 OR iscii-tml)
ISCII,version=6 (aliases: x-iscii-te OR windows-57005 OR iscii-tlg)
ISCII,version=7 (aliases: x-iscii-ka OR windows-57008 OR iscii-knd)

Figure 39: ICU Converter Setup

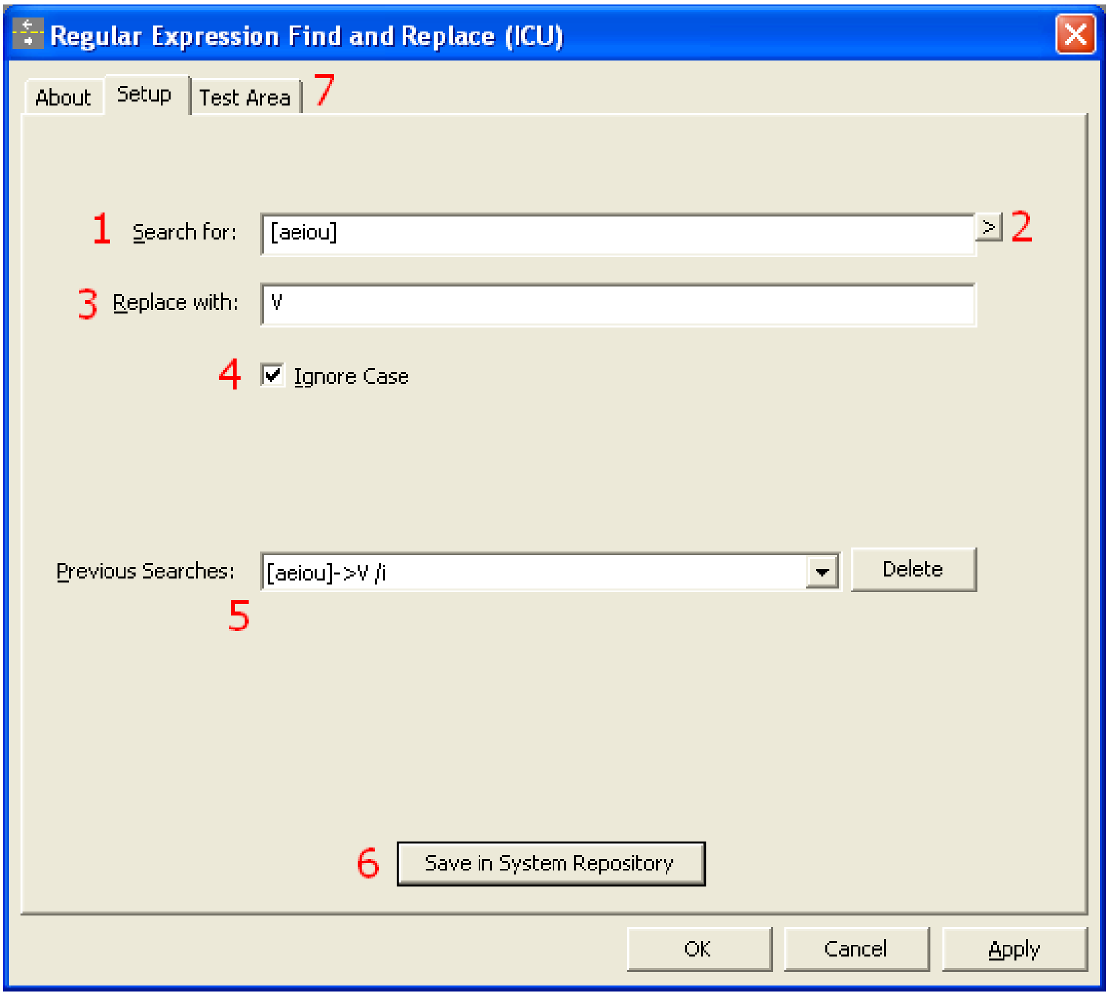
1. Choose the desired converter from the drop-down combo box as shown above.
2. If you want the converter to be permanently added to the System Repository, then you must click Save in System Repository .
3. You can click the **Test Area** tab to test the converter with some sample data.

Regular Expression Find and Replace (ICU)

ICU's Regular Expressions package provides applications with the ability to apply regular expression matching to Unicode string data. The regular expression patterns and behavior are based on Perl's regular expressions. See http://icu.sourceforge.net/userguide/regexp.html for more details on the syntax of ICU Regular Expressions.

- If you want to add an ICU Regular Express Find and Replace converter to the system repository, choose **Regular Expression Find and Replace (ICU)** from the *Choose a Transduction Engine dialog box* and click Add .

The Regular Expression Find and Replace (ICU) Setup dialog will be displayed:



Regular Expression Find and Replace (ICU)

Search for: [aeiou]

Replace with: V

☑ Ignore Case

Previous Searches: [aeiou]->V /i

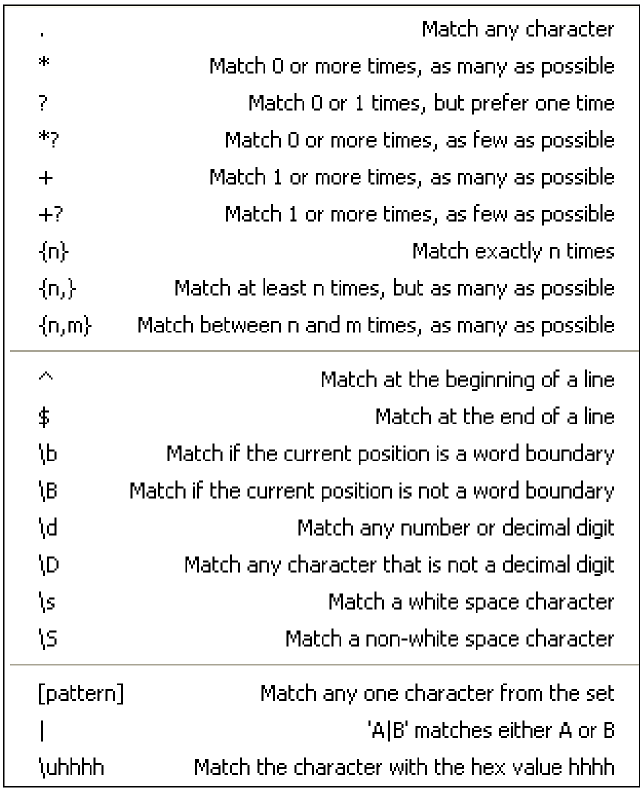
Figure 40: ICU Regular Expression Setup

1. In the **Search for** box, enter the Regular Expression search string you want to use.

   Tip: The search string can contain *Regular Expression Metacharacters* and *Regular Expression Operators* defined below.

2. Click the right wedge button >.

   You will see a pop-up list of commonly used Regular Expression search operators. If selected, they will be inserted into **Search for**.



.	Match any character
*	Match 0 or more times, as many as possible
?	Match 0 or 1 times, but prefer one time
*?	Match 0 or more times, as few as possible
+	Match 1 or more times, as many as possible
+?	Match 1 or more times, as few as possible
{n}	Match exactly n times
{n,}	Match at least n times, but as many as possible
{n,m}	Match between n and m times, as many as possible
^	Match at the beginning of a line
$	Match at the end of a line
\b	Match if the current position is a word boundary
\B	Match if the current position is not a word boundary
\d	Match any number or decimal digit
\D	Match any character that is not a decimal digit
\s	Match a white space character
\S	Match a non-white space character
[pattern]	Match any one character from the set
|	'A|B' matches either A or B
\uhhhh	Match the character with the hex value hhhh

Figure 41: Commonly used regular expressions pop-up

3. In the **Replace with** box, enter the string or operator that represents the text to replace the *Search for* string (see *Replacement Text* defined below).
4. Check the **Ignore Case** box to have the ICU search algorithm ignore the case of the input text.
5. The **Previous Searches** combo box includes a few example Regular Expressions and remembers any new ones you add. Click Delete to remove the selected search item.
6. If you want the converter to be permanently added to the System Repository, then you must click Save in System Repository.
7. Click the **Test Area** tab to test the converter with some sample data.

Regular Expression Metacharacters

Character	Description
\a	Match a BELL, \u0007
\A	Match at the beginning of the input. Differs from ^ in that \A will not match after a new line within the input.
\b, outside of a [Set]	Match if the current position is a word boundary. Boundaries occur at the transitions between word (\w) and non-word (\W) characters, with combining marks ignored. For better word boundaries, see ICU Boundary Analysis at: http://icu.sourceforge.net/userguide/boundaryAnalysis.html
\b, within a [Set]	Match a BACKSPACE, \u0008.
\B	Match if the current position is not a word boundary.
\cX	Match a control-X character.
\d	Match any character with the Unicode General Category of Nd (Number, Decimal Digit.)
\D	Match any character that is not a decimal digit.
\e	Match an ESCAPE, \u001B.
\E	Terminates a \Q ... \E quoted sequence.
\f	Match a FORM FEED, \u000C.
\G	Match if the current position is at the end of the previous match.
\n	Match a LINE FEED, \u000A.
\N{UNICODE CHARACTER NAME}	Match the named character.
\p{UNICODE PROPERTY NAME}	Match any character with the specified Unicode Property.
\P{UNICODE PROPERTY NAME}	Match any character not having the specified Unicode Property.
\Q	Quotes all following characters until \E.
\r	Match a CARRIAGE RETURN, \u000D.
\s	Match a white space character. White space is defined as [\t\n\f\r\p{Z}].
\S	Match a non-white space character.
\t	Match a HORIZONTAL TABULATION, \u0009.
\uhhhh	Match the character with the hex value hhhh.

\Uhhhhhhhh	Match the character with the hex value hhhhhhhh. Exactly eight hex digits must be provided, even though the largest Unicode code point is \U0010ffff.
\w	Match a word character. Word characters are [\p{Ll}\p{Lu}\p{Lt}\p{Lo}\p{Nd}].
\W	Match a non-word character.
\x{hhhh}	Match the character with hex value hhhh. From one to six hex digits may be supplied.
\xhh	Match the character with two digit hex value hh
\X	Match a Grapheme Cluster at: http://www.unicode.org/unicode/reports/tr29/#Grapheme_Cluster_Boundaries
\Z	Match if the current position is at the end of input, but before the final line terminator, if one exists.
\z	Match if the current position is at the end of input.
\n	Back Reference. Match whatever the nth capturing group matched. n must be a number > 1 and < total number of capture groups in the pattern. Note: Octal escapes, such as \012, are not supported in ICU regular expressions
[pattern]	Match any one character from the set. See UnicodeSet at: http://icu.sourceforge.net/userguide/unicodeSet.html for a full description of what may appear in the pattern
.	Match any character.
^	Match at the beginning of a line.
$	Match at the end of a line.
\	Quotes the following character. Characters that must be quoted to be treated as literals are * ? + [( ) { } ^ $ | \ . /

Regular Expression Operators

Operator	Description
|	Alternation. A|B matches either A or B.
*	Match 0 or more times. Match as many times as possible.
+	Match 1 or more times. Match as many times as possible.
?	Match zero or one times. Prefer one.
{n}	Match exactly n times

{n,}	Match at least n times. Match as many times as possible.
{n,m}	Match between n and m times. Match as many times as possible, but not more than m.
*?	Match 0 or more times. Match as few times as possible.
+?	Match 1 or more times. Match as few times as possible.
??	Match zero or one times. Prefer zero.
{n}?	Match exactly n times
{n,}?	Match at least n times, but no more than required for an overall pattern match
{n,m}?	Match between n and m times. Match as few times as possible, but not less than n.
*+	Match 0 or more times. Match as many times as possible when first encountered, do not retry with fewer even if overall match fails (Possessive Match)
++	Match 1 or more times. Possessive match.
?+	Match zero or one times. Possessive match.
{n}+	Match exactly n times
{n,}+	Match at least n times. Possessive Match.
{n,m}+	Match between n and m times. Possessive Match.
(...)	Capturing parentheses. Range of input that matched the parenthesized subexpression is available after the match.
(?: ...)	Non-capturing parentheses. Groups the included pattern, but does not provide capturing of matching text. Somewhat more efficient than capturing parentheses.
(?> ...)	Atomic-match parentheses. First match of the parenthesized subexpression is the only one tried; if it does not lead to an overall pattern match, back up the search for a match to a position before the "(?>"
(?# ...)	Free-format comment (?# comment).
(?= ...)	Look-ahead assertion. True if the parenthesized pattern matches at the current input position, but does not advance the input position.
(?! ...)	Negative look-ahead assertion. True if the parenthesized pattern does not match

	at the current input position. Does not advance the input position.
(?<= ...)	Look-behind assertion. True if the parenthesized pattern matches text preceding the current input position, with the last character of the match being the input character just before the current position. Does not alter the input position. The length of possible strings matched by the look-behind pattern must not be unbounded (no * or + operators.)
(?<! ...)	Negative Look-behind assertion. True if the parenthesized pattern does not match text preceding the current input position, with the last character of the match being the input character just before the current position. Does not alter the input position. The length of possible strings matched by the look-behind pattern must not be unbounded (no * or + operators.)
(?ismx-ismx: ...)	Flag settings. Evaluate the parenthesized expression with the specified flags enabled or -disabled.
(?ismx-ismx)	Flag settings. Change the flag settings. Changes apply to the portion of the pattern following the setting. For example, (?i) changes to a case insensitive match.

Replacement Text

The replacement text for find-and-replace operations may contain references to capture-group text from the find. References are of the form $n, where n is the number of the capture group.

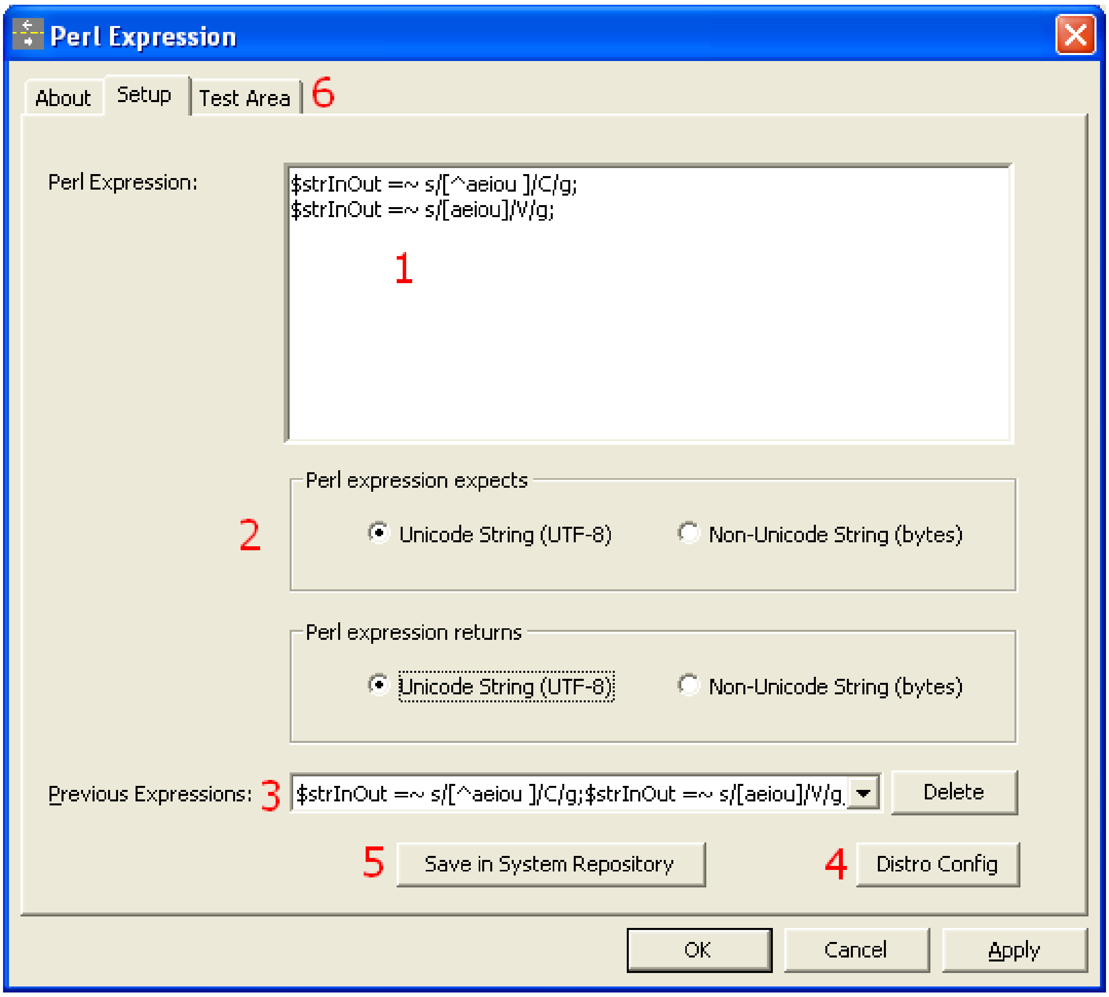
Character	Descriptions
$n	The text of capture group n will be substituted for $n. n must be >= 0 and not greater than the number of capture groups. A $ not followed by a digit has no special meaning, and will appear in the substitution text as itself, a $.
\	Treat the following character as a literal, suppressing any special meaning. Backslash escaping in substitution text is only required for '$' and '\', but may be used on any other character without bad effects.

Perl Expression

> This feature requires installation of a separate Perl 5.10.1 distribution to be installed.
>
> The Perl plug-in has been tested with the freely available Perl distribution from ActiveState at: http://www.activestate.com/activeperl.

To add a Perl Expression converter to the system repository, choose **Perl Expression** from the *Choose a Transduction Engine dialog box* and click Add .

The Perl Expression Setup dialog will be displayed:



Perl Expression:

```
$strInOut =~ s/[^aeiou ]/C/g;
$strInOut =~ s/[aeiou]/V/g;
```

Perl expression expects: ● Unicode String (UTF-8) ○ Non-Unicode String (bytes)

Perl expression returns: ● Unicode String (UTF-8) ○ Non-Unicode String (bytes)

Previous Expressions: $strInOut =~ s/[^aeiou]/C/g;$strInOut =~ s/[aeiou]/V/g

Figure 42: Perl Expression Setup

1. This area is for entering your Perl expression.
2. For *Perl expression expects* and *Perl expression returns*, select the desired encoding type and click Apply . *Tip*: If your data is Unicode-encoded, select that option or your data may be incorrectly converted. For *Non-Unicode (byte) data*, the default system code page will be used to convert your data when necessary.
3. The **Previous Expressions** combo box includes a few example Perl expressions and remembers any new ones you add. Click Delete to remove unwanted expressions.
4. Click Distro Config to set up the path to your Perl Distribution's library folders (e.g. C:\Perl, C:\Perl\lib, C:\Perl\site\lib) and to specify certain Perl modules to be automatically loaded for all expressions (e.g. Win32, or SIL:RTF:Unicode). *Tip*: If you get an error message saying that some *.pm file can't be found, then you probably don't have the correct path to the file configured.
5. To permanently add the expression to the System Repository, click Save in System Repository .

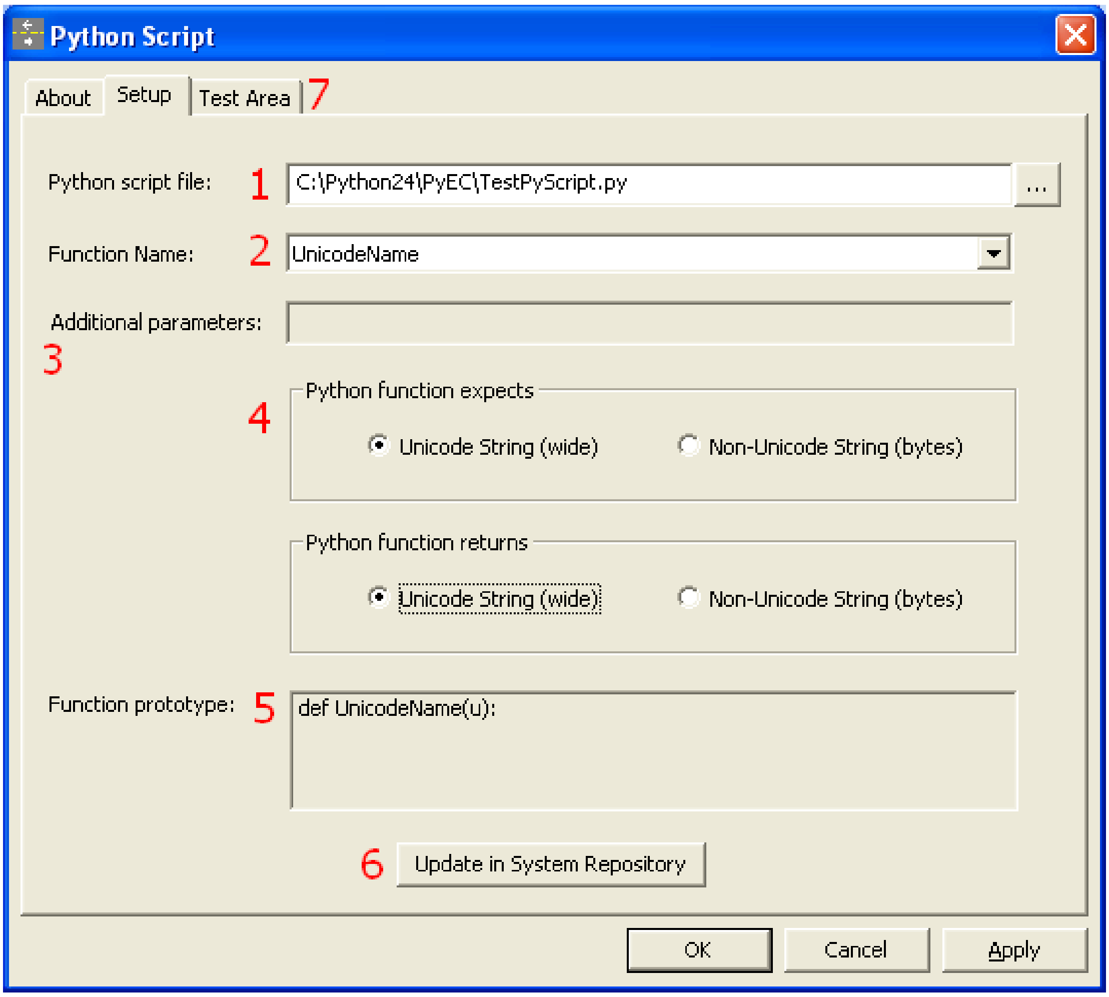
6. Click the **Test Area** tab to test the converter with some sample data.

Python Script Function

> This feature requires a separate Python 2.5 distribution to be installed.
>
> The Python plug-in has been tested with the following freely available Python distributions: ActiveState Python at: http://www.activestate.com/solutions/perl or Python.org at: http://www.python.org/download.

If you want to add a Python script function converter to the system repository, choose **Python Script** from the *Choose a Transduction Engine dialog box* and click Add .

The Python Script Setup dialog will be displayed:



Python script file: C:\Python24\PyEC\TestPyScript.py

Function Name: UnicodeName

Additional parameters:

Python function expects: ● Unicode String (wide) ○ Non-Unicode String (bytes)

Python function returns: ● Unicode String (wide) ○ Non-Unicode String (bytes)

Function prototype: def UnicodeName(u):

Figure 43: Python Script Setup

1. Browse with the ... button for the Python script file.
2. If the file contains valid Python functions, then the function names will be put into the **Function Name** combo box. Choose the function desired.

3. If you want to pass some static information to your Python function, then define the function as shown below and put the static information in the **Additional parameters** field. For example, with a Python function prototype of:

```
def ChangeLanguage(sLang, uI):
    if not isinstance(uI, unicode):
        raise UnicodeError(u'Input Data not Unicode! (%s)' % uI)0
    else:
        if sLang == u'Chinese':
            # do some Chinese processing and put result in uO
            uO = ProcessChinese(uI)
        return uO
```

   The **Additional parameters** field would be enabled and you could enter the fixed string, **Chinese**, in order to trigger the script properly.

> If you have more than one additional parameter, the static strings should be separated by a semicolon (i.e. ";").

4. This area is for indicating how your expression expects and returns data. If your data is Unicode-encoded, be sure to select that option or your data may be incorrectly converted. For *Non-Unicode (byte) data*, the default system code page will be used to convert your data when necessary.

5. After configuring these four items, click Apply to accept the Python script function.

   The Setup tab also has the following options:

   - As you iterate through the functions listed in the combo box in, the Function Prototype window show what the prototype of selected Python function looks like.

> If a particular function allows additional (static) parameters, then the proper order of parameters will also be shown in this window.

6. If you want the Python function to be permanently added to the System Repository, then you must click Save in System Repository .

7. Click the **Test Area** tab to test the converter with some sample data.

AdaptIt Knowledge Base Lookup Converter

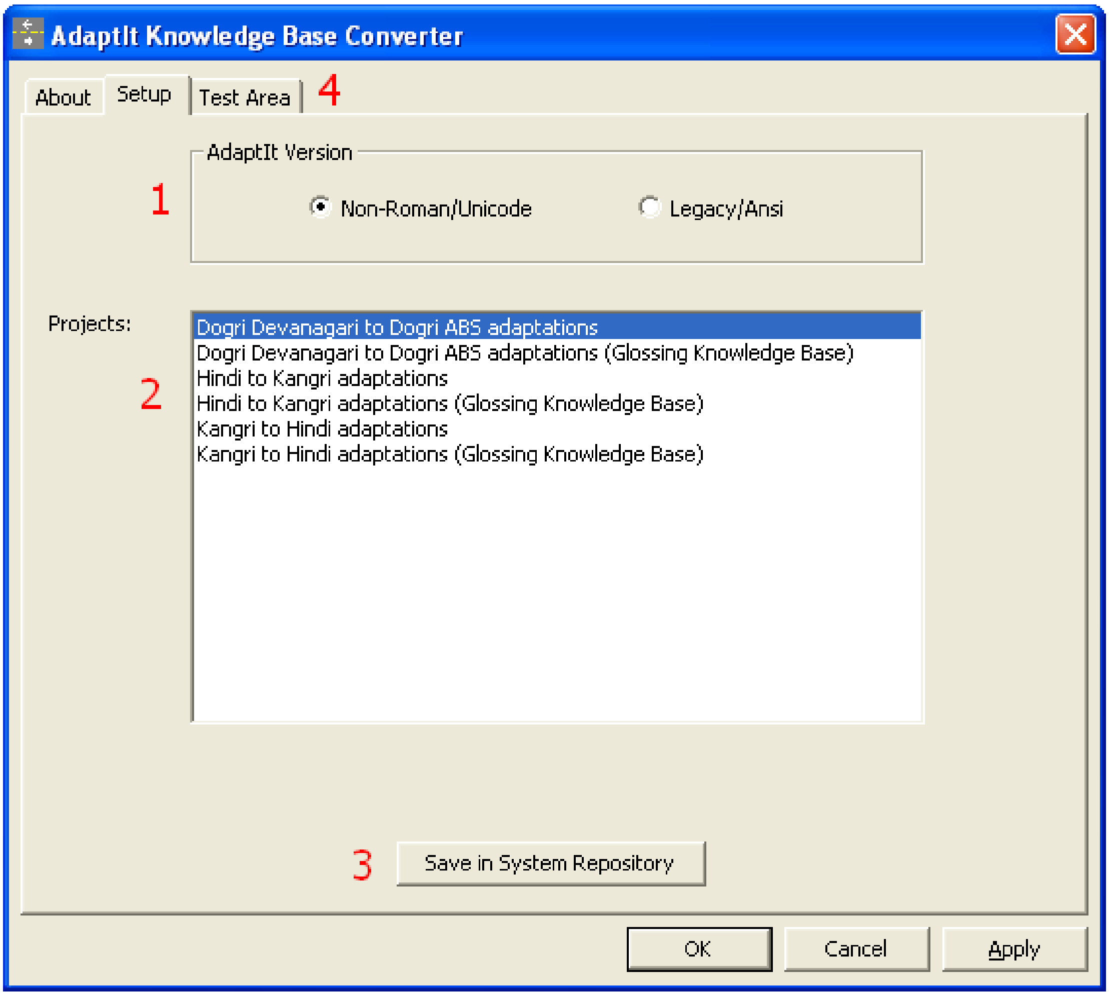
This transducer allows you to do lookups on words in either the *adaptation* or *glossing* Knowledge Base of an AdaptIt project.

> The *AdaptIt Knowledge Base Lookup* transduction engine is installed as part of the Adapt It transduction engine package in the SILConverters 3.1 installer and will not be visible in the Choose Transduction Engine dialog box unless you have selected that option for installation (which is not installed by default).

To add an AdaptIt Knowledge Base Lookup converter to the system repository:

- Choose **AdaptIt Knowledge Base Converter** from the *Choose a Transduction Engine dialog box* and click Add .

The AdaptIt Knowledge Base Converter Setup dialog will be displayed:



AdaptIt Version

● Non-Roman/Unicode ○ Legacy/Ansi

Projects:

Dogri Devanagari to Dogri ABS adaptations
Dogri Devanagari to Dogri ABS adaptations (Glossing Knowledge Base)
Hindi to Kangri adaptations
Hindi to Kangri adaptations (Glossing Knowledge Base)
Kangri to Hindi adaptations
Kangri to Hindi adaptations (Glossing Knowledge Base)

Figure 44: AdaptIt Knowledge Base Lookup Converter Setup

1. Choose which version of AdaptIt (i.e. **Non-Roman/Unicode** vs. **Legacy/Ansi**) that you use to create the (XML) Knowledge Base.
2. Choose the project desired in the **Projects** list box. This is automatically populated from the projects available on the local machine for the current user.

> For an AdaptIt *Transliteration* Project, the transliteration data will be in the normal project Knowledge Base file; *not* the *Glossing Knowledge Base*. However, it is possible to access a Glossing Knowledge Base if desired.

> Note that if you access an *Adaptation* Knowledge Base (i.e. from an AdaptIt project used to adapt texts from one language to another—which most likely will contain ambiguities), then

the converter will return a string containing all the ambiguities for the given lookup word in the *Ample ambiguity format* (i.e. %count%form1%form2%...%).

For example, if your AdaptIt Knowledge Base has an ambiguity for the word /से/ 'from, with' in the *Source* language, which sometimes means /ते/ 'from' and sometimes /कन्ने/ 'with' in the target language, then if you attempt to process the word /से/ with this converter, it will return the string /%2%ते%कन्ने%/.

If you have such values in a document readable by Microsoft Word, then you can use the *Word Pick* Document Template to simplify disambiguating these tokens.

3. If you want the converter to be permanently added to the System Repository, then you must click Save in System Repository .
4. Click **Test Area** tab to test the converter with some sample data.

AdaptIt Target Word Guesser

This transducer can be used in an Adapt It adaptation project to guess target words based on the Source/Target word pairs in the Adapt It Knowledge Base.

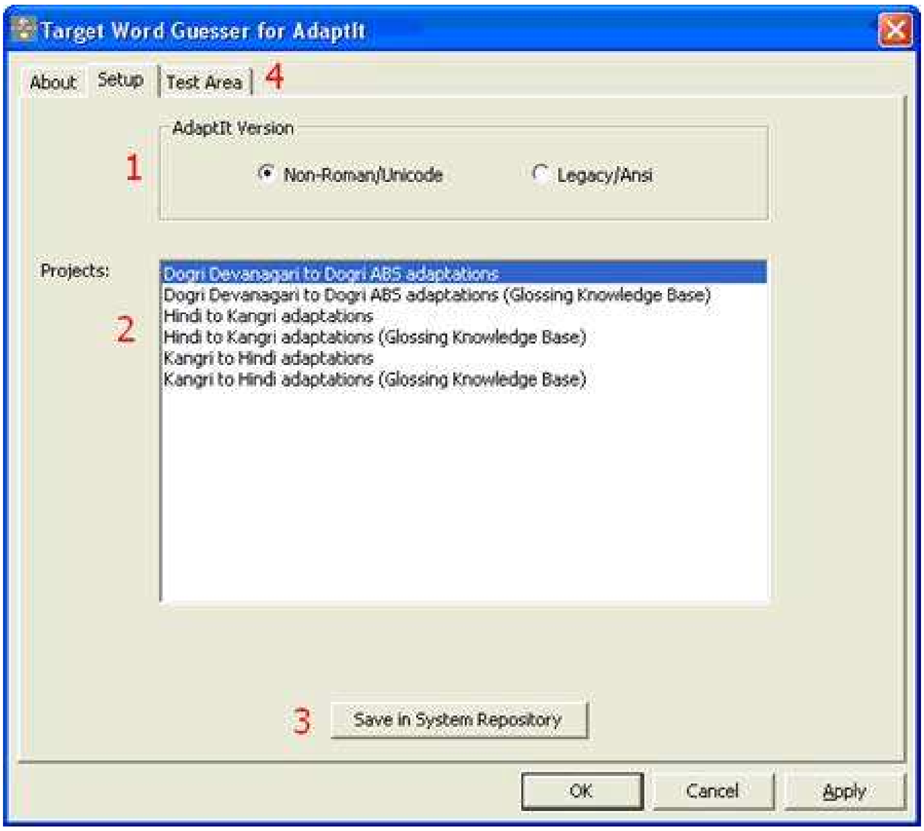
Forthcoming versions of Adapt It will have this functionality built-in, but until then (roughly Fall ’07), this functionality is available thru this Target Word Guesser for AdaptIt EncConverter plug-in.

The *AdaptIt Knowledge Base Lookup* transduction engine is installed as part of the Adapt It transduction engine package in the SILConverters 3.1 installer and will not be visible in the Choose Transduction Engine dialog box unless you have selected that option for installation (which is not installed by default).

To add an AdaptIt Target Word Guesser transducer to the system repository:

- Choose **Target Word Guesser for AdaptIt** from the *Choose a Transduction Engine dialog box* and click Add .

The AdaptIt Target Word Guesser Setup dialog will be displayed:



AdaptIt Version

● Non-Roman/Unicode ○ Legacy/Ansi

Projects:

Dogri Devanagari to Dogri ABS adaptations
Dogri Devanagari to Dogri ABS adaptations (Glossing Knowledge Base)
Hindi to Kangri adaptations
Hindi to Kangri adaptations (Glossing Knowledge Base)
Kangri to Hindi adaptations
Kangri to Hindi adaptations (Glossing Knowledge Base)

Figure 45: AdaptIt Knowledge Base Lookup Converter Setup

1. Choose which version of AdaptIt (i.e. **Non-Roman/Unicode** vs. **Legacy/Ansi**) that you use to create the (XML) Knowledge Base.
2. Choose the project desired in the **Projects** list box. This is automatically populated from the projects available on the local machine for the current user. You will want to choose the same project for which the guesser is to be used (see below).
3. If you want the converter to be permanently added to the System Repository, then you must click the **Save in System Repository** button (see below).
4. You can click on the **Test Area** tab in order to test the converter with some sample data.

   For example, given an AdaptIt knowledge base with the following word pairs:

Source	Target
boyes	boyez
girles	girlez
guyes	guyez

Then for the source word pearles, the guesser will return pearlez.

Compound "daisy-chained" converters

Two final converter types to be discussed are actually *meta converter* types; that is, they allow you to combine two or more existing converters in the system repository in a serial or parallel fashion.

The *Compound Converter* type can be used to combine 2 or more converters together in a serial fashion so that the output of one step will become the input to the next step automatically. This can be helpful when you have multiple, different conversions to apply to your data to get it in the ultimate form you need without requiring separate conversions.

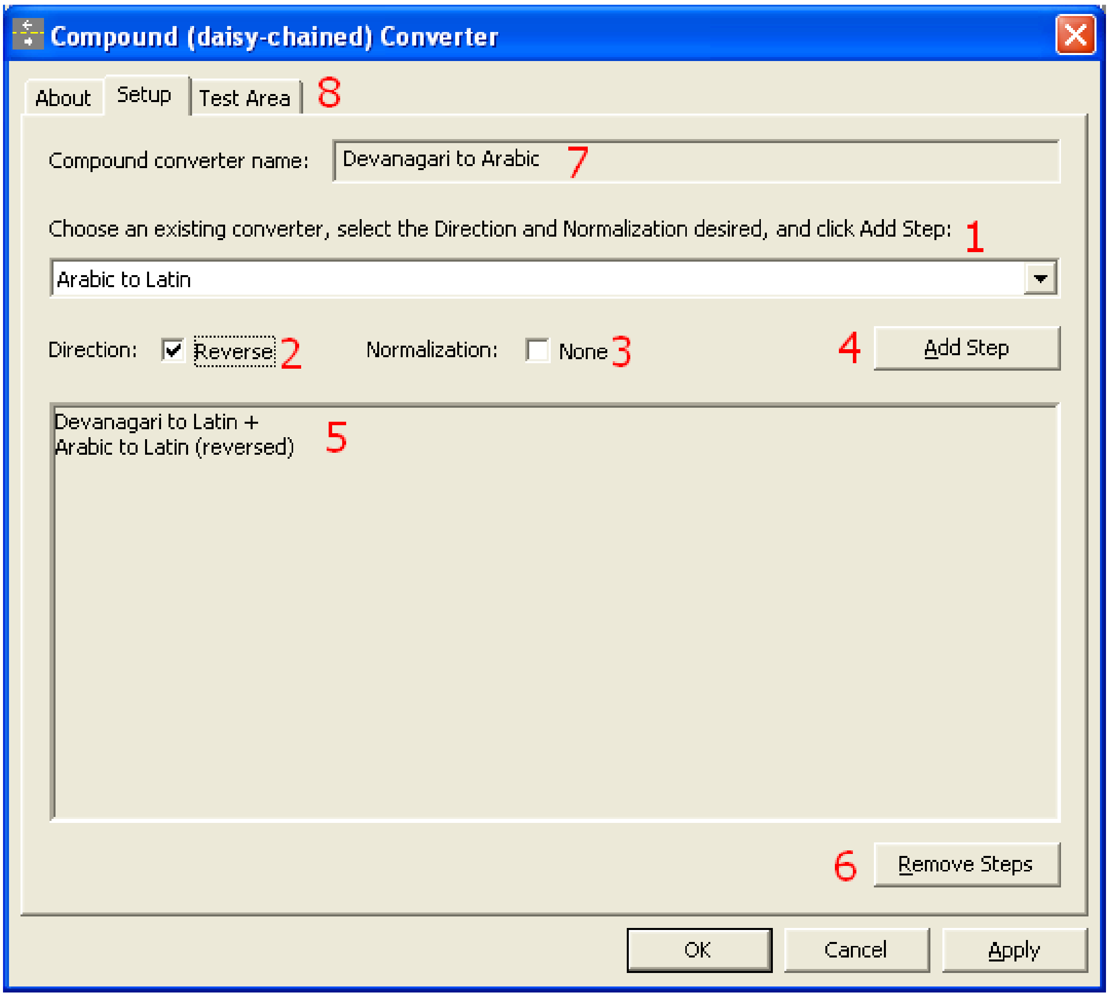
For example, you may have one converter that goes from FindPhone IPA to SIL-IPA93, and another converter that converts from SIL-IPA93 to Unicode IPA. In order to perform the end-to-end conversion from FindPhone IPA to Unicode IPA, you can create a daisy-chain of the two existing converters (a "virtual converter") so that the data is converted in one step.

> When creating or using a compound converter, then all n+1 converters must be in the system repository (i.e. the n steps plus the compound converter itself). If you create a compound converter and then subsequently delete one of the steps, it will not work.

To add a Compound converter to the system repository:

- Choose **Compound (daisy-chained) Converter** from the *Choose a Transduction Engine dialog box* and click Add .

The Compound (daisy-chained) Converter Setup dialog will be displayed:



Compound (daisy-chained) Converter
About
Setup
Test Area
8
Compound converter name:
7
Choose an existing converter, select the Direction and Normalization desired, and click Add Step:
1
Direction:
2
Normalization:
3
4
Add Step
5
6
Remove Steps
OK
Cancel
Apply

Compound converter name: Devanagari to Arabic

Arabic to Latin

Direction: ☑ Reverse

Normalization: □ None

Devanagari to Latin +
Arabic to Latin (reversed)

Figure 46: Compound Converter Setup

1. Use this combo box to choose the converter to become one of the *steps* of the compound conversion.
2. If the selected converter is bi-directional, then the **Reverse** box will be enabled allowing you to select the reverse direction, if needed. For example, to go from Devanagari to Arabic, note that the above configured converter goes *forward* (by default) from Devanagari to Latin. This is followed by a second step that goes in the *reverse* direction from Latin to Arabic.
3. If you need to explicitly normalize the output data of any step (i.e. to *Fully Composed* or *Full Decomposed*), you can check the **Normalization** box and the compound converter will do this before continuing with the next step.
4. Click Add Step to add it to the queue of steps in the compound converter.
5. This area shows what steps are configured, the direction of conversion, and whether any normalization is requested.
6. If you make a mistake in the steps, click Remove Steps to clear out the compound converter and start over.

Once you have configured all the steps, you click Apply to save the converter in the System Repository.

> Compound converters may not be temporary converters.

Once you click Apply, the **Enter Converter Name** dialog box will appear

> The converter friendly name you enter here is for the *Compound converter* itself, which is distinct from the names of the converter steps. For Compound converters, the default name will a concatenation of the individual steps' names. However, you can change it to something more meaningful if desired (e.g. *Devanagari to Arabic*).

7. Once you complete the previous step, the converter name will be displayed in the **Compound converter name** box.
8. Click the **Test Area** tab to test the converter with some sample data.

Primary-Fallback Compound Converter

The *Primary-Fallback Compound Converter* type allows you to specify two existing converters: one to be a primary, and the other, a fallback converter.

The configured primary converter is first called to do a conversion. If the primary converter doesn't change the data, then, and only then, the fallback converter is called.

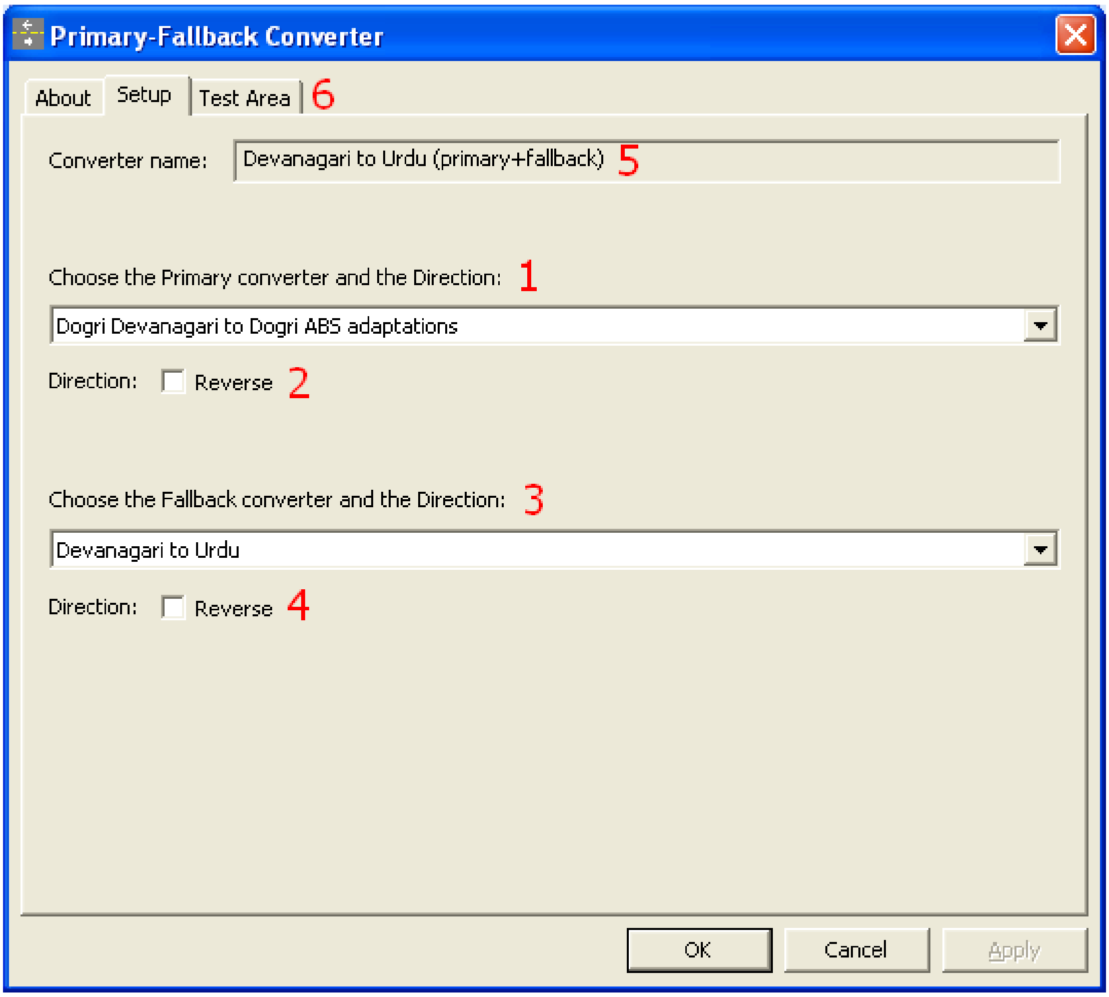
This can be useful for transliteration where a character-based transliterator (e.g. TECkit, ICU, or CC) does most of the work, but certain words (or character sequences) are otherwise unpredictable from the context. In this case, you might want a lexicon-based approach to supply the special case forms.

In this scenario, you would configure the lexicon-based transliterator (e.g. a *SpellFixer* CC table or an *AdaptIt Knowledge Base Lookup* converter) to be the primary converter and the character-based transliterator as the fallback converter. If the text isn't modified by the primary converter (i.e. if it isn't an exception), then the fallback (algorithmic) converter is called to do the conversion.

To add a Primary-Fallback converter to the system repository:

- Choose **Primary-Fallback Converter** from the *Choose a Transduction Engine dialog box* and click Add.

The Primary-Fallback Converter Setup dialog will be displayed:



Converter name: Devanagari to Urdu (primary+fallback)

Choose the Primary converter and the Direction:

Dogri Devanagari to Dogri ABS adaptations

Direction: □ Reverse

Choose the Fallback converter and the Direction:

Devanagari to Urdu

Direction: □ Reverse

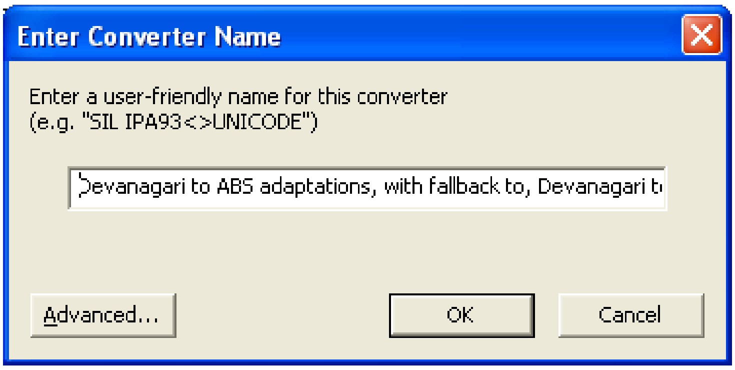
Figure 47: Primary-Fallback Converter Setup

1. Use this combo box to choose the existing EncConverter which is to be the *Primary* converter for this conversion (e.g. the lexical lookup converter).
2. If the selected Primary converter is bi-directional, then the Reverse box will be enabled allowing you to select the reverse direction, if needed.
3. Use this combo box to choose the existing EncConverter which is to be the Fallback converter for this conversion (e.g. the character-based transliterator).
4. If the selected Fallback converter is bi-directional, then the Reverse box will be enabled allowing you to select the reverse direction, if needed.

   Once you have configured both the Primary and Fallback converters, you click Apply to Save the converter in the System Repository.

Primary-Fallback converters may not be temporary converters.

Once you click Apply, the **Enter Converter Name** dialog box will appear:



Enter Converter Name

Enter a user-friendly name for this converter (e.g. "SIL IPA93<>UNICODE")

Devanagari to ABS adaptations, with fallback to, Devanagari t

Figure 48: Enter Converter Name dialog

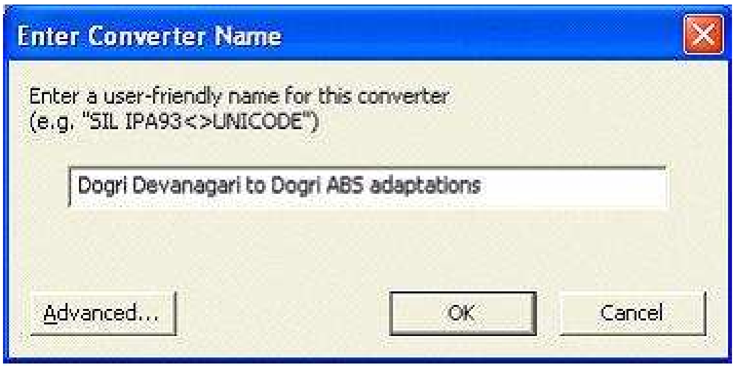
The converter friendly name you enter here is for the combined *Primary-Fallback* converter itself, which is distinct from the names of the Primary and Fallback converters.

5. Once you complete the previous step, the converter name will be displayed in the **Compound converter name** box.
6. Click the **Test Area** tab to test the converter with some sample data.

Saving the converter in the System Repository

On any of the Transduction Engine Setup dialogs, by default, if you click Apply or OK, the configured converter will be returned to the client application as a temporary converter; once the client application (e.g. FieldWorks or Word) is closed or releases the converter, it will no longer be available. If you want the converter to be permanently available to client applications, then you must explicitly add it to the System Repository using the Save in System Repository button (or the Update in System Repository button when editing a map).

- When you click Save in System Repository, the following dialog box will be displayed to query for a friendly name by which the converter will be known in client applications:



Enter Converter Name

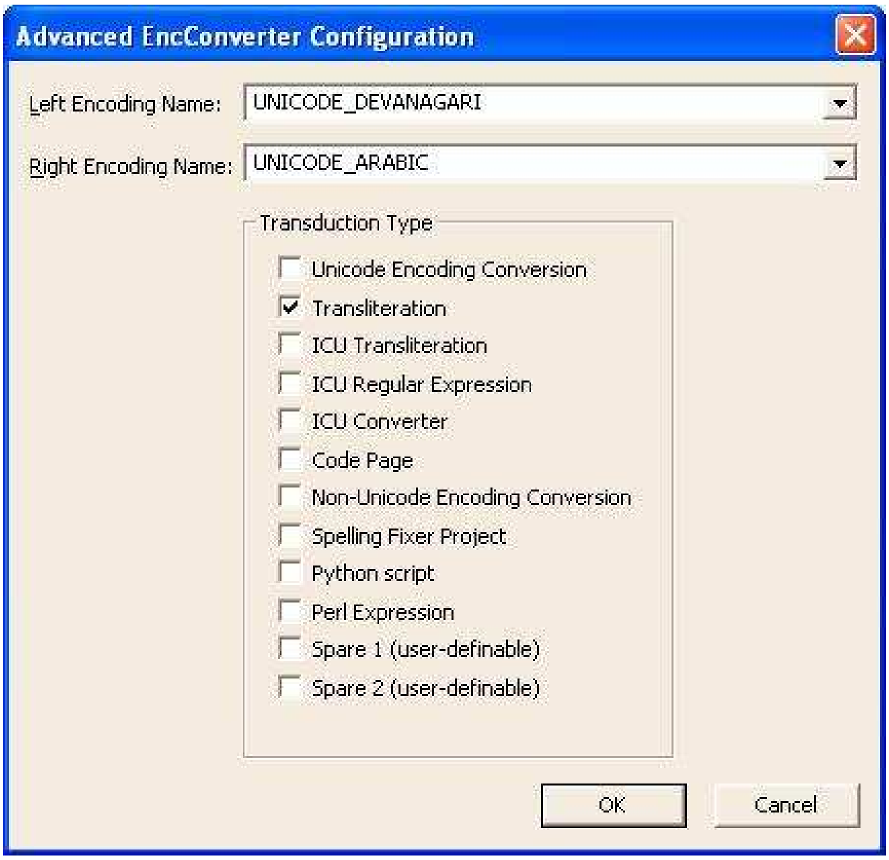
Enter a user-friendly name for this converter (e.g. "SIL IPA93<>UNICODE")

Dogri Devanagari to Dogri ABS adaptations

Figure 49: Enter Converter Name dialog box

- Click Advanced... to enter further, optional information about this converter which is also put into the System Repository.

  The **Advanced EncConverter Configuration** dialog will be displayed:



Left Encoding Name: UNICODE_DEVANAGARI

Right Encoding Name: UNICODE_ARABIC

Transduction Type

- □ Unicode Encoding Conversion
- ☑ Transliteration
- □ ICU Transliteration
- □ ICU Regular Expression
- □ ICU Converter
- □ Code Page
- □ Non-Unicode Encoding Conversion
- □ Spelling Fixer Project
- □ Python script
- □ Perl Expression
- □ Spare 1 (user-definable)
- □ Spare 2 (user-definable)

Figure 50: Advanced Configuration dialog box

Though these values are not necessary for the operation of the converter, they can be helpful to various client applications. For example, the Clipboard EncConverter can be configured to filter the list of displayed converters based on the Encoding Name and/or the Transduction Type configured here.