

Contents

What is GumPad?.....	1
The GumPad Editor.....	1
Convert menu	2
Examples	3
Preferences menu.....	4
Fonts.....	4
Map	4
Convert As You Type	8
Recognized file extensions.....	9
Add-In for Microsoft Word	10

What is GumPad?

GumPad is a text processor that converts text entered in English into text in various Indian languages and back to English or into another Indian language. It can be used to create, edit, view, and print simple text documents in Devanagari, Telugu, Tamil, Kannada, Malayalam, Marathi, Gujarati, Gurmukhi, Bengali and Oriya. It is published as a standalone editor as well as an Add-In for Microsoft Word 2007 and Word 2010.

The GumPad Editor

The editor is similar to *Notepad* on Windows. It has a document area where you type text, make changes, and convert text between English, Indian languages and Extended Latin. There is a menu bar at the top of this document area. The menu bar has a Convert menu that is used to convert text and a Preferences menu to set conversion preferences.

Convert menu

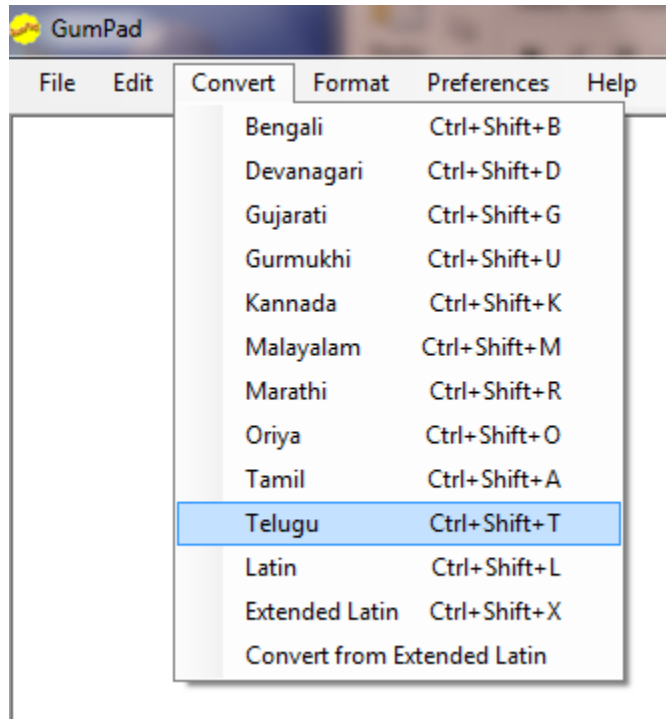


FIGURE 1 THE CONVERT MENU

After typing some text in English, click the Convert menu and select a language to convert it into the selected language. If you would like to convert only a portion of the text in the document, select it first and then use the convert menu.

Tip

Most menu actions can also be performed via keyboard shortcuts. Shortcuts are listed alongside menu items in each menu. For e.g., **Ctrl+Shift+T** is the keyboard shortcut to convert text to Telugu. To use this shortcut, type or select some text and press the **Control**, **Shift** and **'T'** keys together at the same time. The Convert menu is also available as a context menu from the document window. Click the right mouse

button in the document area to view the context menu.

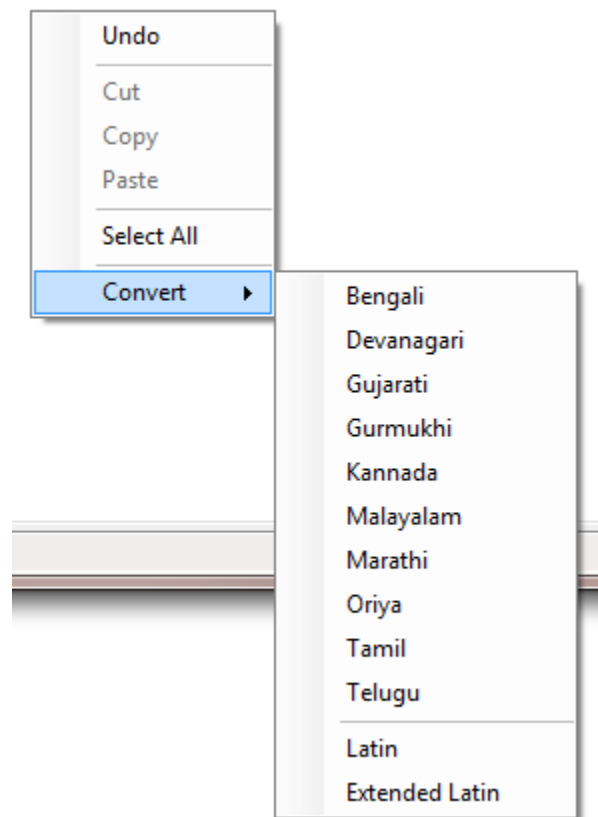


FIGURE 2 THE CONTEXT MENU

Note

GumPad supports two conversion modes – “Convert as you type” and “Convert after you type”. The Convert menu is visible in the menu bar only in the “Convert after you type” mode. You can switch between the two modes via the “Convert as you type” checkbox on the Preferences menu.

Examples

1. Type **vandE maataram** (English) followed by **Ctrl+Shift+D** to convert it to **वन्दे मातरम्** (Devanagari) followed by **Ctrl+Shift+X** to convert it into **vandē mātaram** (Extended Latin).
2. Type **sau.ndarya lahari** followed by **Ctrl+Shift+T** to convert it to **ಸೌಂದರ್ಯ ಲಹರಿ** followed by **Ctrl+Shift+L** to convert it back into **sau.ndarya lahari**.

“Convert from Extended Latin checkbox” in the Convert menu

Leave this unchecked for normal conversion from English to other Indian languages or when converting Indian language text back to English. Leave it checked when you want to convert Extended English text to an Indian language. Extended English text is English text with diacritics.

Note

GumPad cannot convert text directly from English to Extended English. Convert text from English to an Indian language and then convert the text in the Indian language to Extended English.

Example

Type **himAlaya** followed by **Ctrl+Shift+D** to convert it to **हिमालय** followed by **Ctrl+Shift+X** to convert it into **himālaya**. Check the **Convert from Extended Latin** checkbox on the Preferences menu followed by **Ctrl+Shift+D** to convert **himālaya** to **हिमालय**

Preferences menu

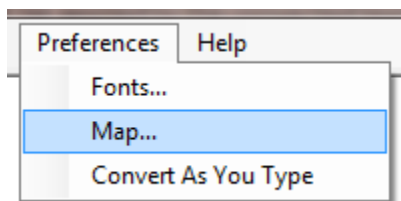


FIGURE 3 THE PREFERENCES MENU

Fonts

Use this menu item to set the fonts for converted text. The built-in settings are based on fonts that ship with Windows.

Map

GumPad uses a transliteration map to convert sequences of English characters into sequences of Indian characters. A unique feature of GumPad is the option to change this conversion map to suit your own purposes. In normal usage, there is no need to change the conversion map. However, there are several transliterators available for Indian languages and a few popular ones follow a slightly different transliteration scheme.

This option is useful for converting text encoded in a scheme that is different from the one provided by GumPad. Most of GumPad's built-in mappings are based on ITRANS (<http://www.aczoom.com/itrans/>) with a few differences such as "e" and "o" to allow for stretched versions of these vowels in other Indian languages.

Select the *Map...* menu item from the Preferences menu to change the conversion map.

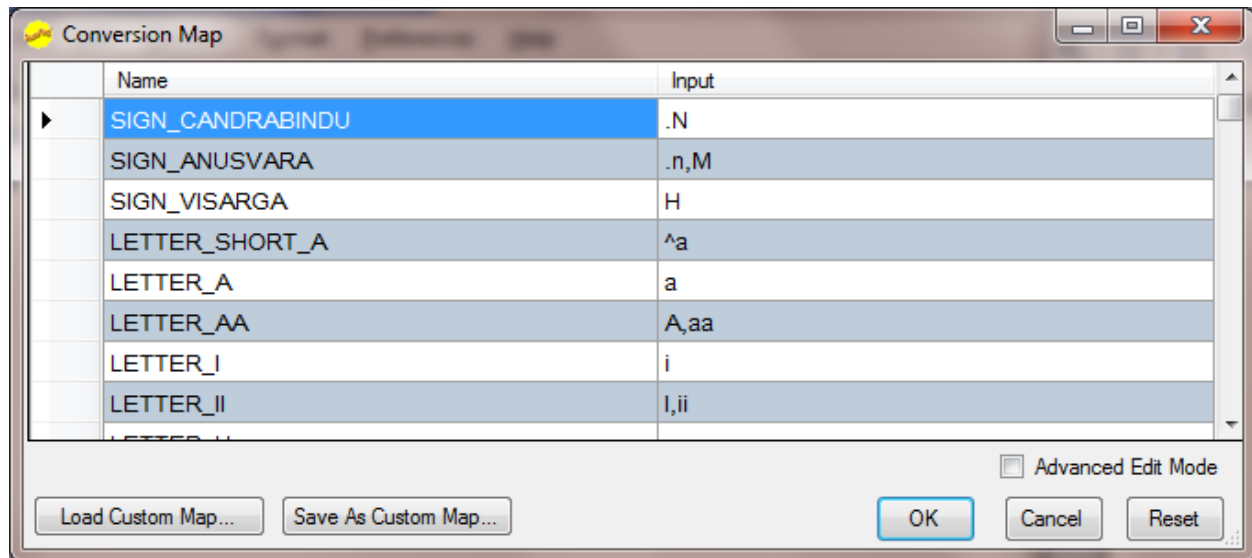


FIGURE 4 CONVERSION MAP - SIMPLE EDIT MODE

You should see a table listing the names of Indian letters corresponding to the Input English text for each letter. The *Name* column lists the name of the Indian character or character sequence that will be output when the text in the *Input* column is typed. For simple changes, click on the Input column of a letter and type in a new sequence. Click the *OK* button to commit your changes when you are done. Separate Input sequences using commas if you want more than one sequence of input letters to map to the same output sequence. Click *Cancel* to abandon changes or *Reset* to revert to the built-in map at any time.

Example

'a' converts to अ while 'aa' or 'A' converts to आ. This is shown in the table as 'a' for LETTER_A and 'A,aa' for LETTER_AA.

Note

Input sequences must be unique. i.e., you cannot map the same sequence to produce more than one output sequence. An error message will be displayed if a duplicate mapping is detected.

Click on the *Advanced Edit Mode* checkbox to view and use a more detailed version of the map editor.

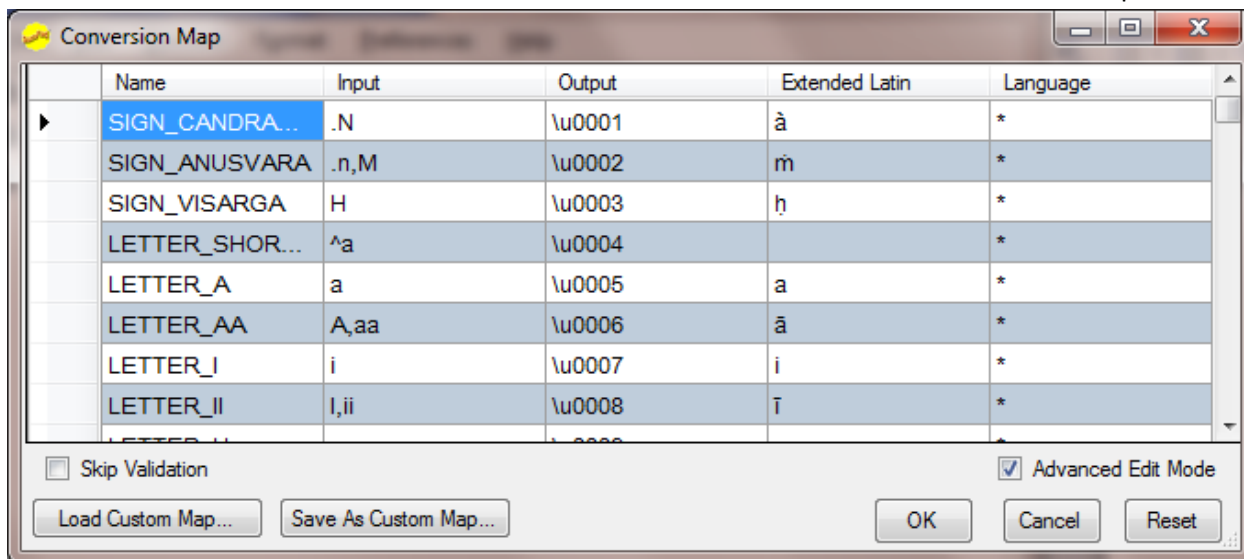


FIGURE 5 CONVERSION MAP - ADVANCED EDIT MODE

In this mode, you can change not only the Input sequence in English, but also the output sequence of Unicode characters and Extended Latin characters.

Note

The Unicode output sequence is a sequence of characters in the format \uNNNN where NNNN corresponds to a Unicode code point. For e.g., SIGN_OM corresponds to Unicode U+0950 and is entered as \u0950. There is a common block of Indian vowels and consonants that have to be entered as \u00NN where 00NN corresponds to the offset of the common block character from the starting point of the Unicode code chart for that language. GumPad will use the offset to correctly output the final Unicode character for the language it is converting to. For e.g., LETTER_A is represented as \u0005 in GumPad's Unicode output. LETTER_A being a common block character, will be output as U+0905 in Devanagari or U+0C05 in Telugu. This is the core mechanism that GumPad leverages to transliterate common block characters between different Indian languages. Unicode also assigns U+FFFE and U+FFFF as Non-characters. Use either of these in an output sequence to suppress output for the corresponding input sequence. Unicode code charts for all languages are available at <http://unicode.org/>.

Conjuncts

Certain letter sequences go together and are treated as conjuncts. CONJUNCT_KSHA is a familiar example. The default input sequence for this conjunct is *ksh* or *kSh* or *x*. The default output sequence is \u0015\u004D\u0037, i.e. LETTER_KA followed by SIGN_VIRAMA (Halant) followed by LETTER_SHA.

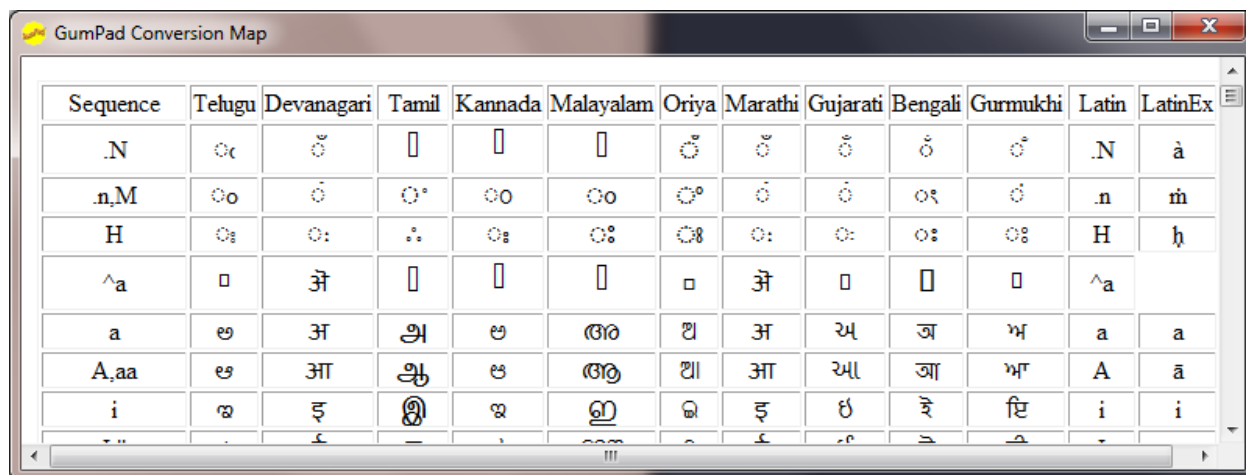
The Language Column

Use this column for output sequences that are specific to a particular language. For e.g., CONJUNCT_KKA is specific to Devanagari and is associated with "hi", the ISO language identifier for Hindi whereas LETTER_A is common to all Indian languages and is associated with a "*" in the language column.

Load and Save Custom Maps

A customized map can be saved to an external .map file via the *Save As Custom Map* button. Saved maps can be shared or reloaded via the *Load Custom Map* button.

Select the *Show Map...* menu item from the *Help* menu to view the current map.



Sequence	Telugu	Devanagari	Tamil	Kannada	Malayalam	Oriya	Marathi	Gujarati	Bengali	Gurmukhi	Latin	LatinEx
.N	౯	ँ	஀	஀	஀	ଌ	ँ	ँ	ঁ	ँ	.N	à
.n,M	ం	ं	஁	ಂ	ం	ୠ	ं	ं	ৎ	ं	.n	m
H	ః	ः	ஃ	ః	ః	ଌ	ः	ः	ঃ	ః	H	h
^a	౱	अ	஀	஀	஀	ଐ	अ	౱	ঐ	ਐ	^a	
a	అ	अ	அ	అ	അ	ଅ	अ	અ	অ	ਅ	a	a
A,aa	ఆ	आ	ஆ	ಆ	ആ	ଌ	आ	અ	আ	ਆ	A	ā
i	ఇ	इ	இ	ಇ	ഇ	ଈ	इ	ઇ	ই	ਇ	i	i

FIGURE 6 VIEW CONVERSION MAP

Custom Maps

Custom maps are XML files with a .map extension. These can be modified in any editor and loaded into GumPad. The XML schema is fairly straightforward. There is a <TransliterationMap> root element within which are several <mapping> elements. Each <mapping> element corresponds to a mapping rule for a single output character or character sequence. Here is an example that shows the mappings for LETTER_A, LETTER_AA and CONJUNCT_KSHA. Notice how LETTER_AA and CONJUNCT_KSHA have more than one input sequence whereas LETTER_A has only one. Notice also how the output for CONJUNCT_KSHA is defined as a Unicode sequence as opposed to a single character for LETTER_A or LETTER_AA. Study the built-in map and the sample map for ITRANS for more examples. This is a powerful mechanism to add to, modify or even completely replace the built-in transliteration map.

```
<TransliterationMap xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xsd="http://www.w3.org/2001/XMLSchema" schema-revision="2.0"
transliteration-scheme-name="ITRANS"
contributor-name="GumPad sample conversion map for ITRANS">
  <mapping>
    <character-name>LETTER_A</character-name>
    <input-sequence>a</input-sequence>
    <unicode-output-sequence>\u0005</unicode-output-sequence>
    <extended-latin-output-sequence>a</extended-latin-output-sequence>
    <language>*</language>
  </mapping>
  <mapping>
    <character-name>LETTER_AA</character-name>
```

```
<input-sequence>A</input-sequence>
<input-sequence>aa</input-sequence>
<unicode-output-sequence>\u0006</unicode-output-sequence>
<extended-latin-output-sequence>ā</extended-latin-output-sequence>
<language>*</language>
</mapping>
<mapping>
  <character-name>CONJUNCT_KSHA</character-name>
  <input-sequence>kSh</input-sequence>
  <input-sequence>ksh</input-sequence>
  <input-sequence>x</input-sequence>
  <unicode-output-sequence>\u0015\u004D\u0037</unicode-output-sequence>
  <extended-latin-output-sequence />
  <language>*</language>
</mapping>
</TransliterationMap>
```

Convert As You Type

Use this checkbox to switch between convert as you type and convert after you type modes.

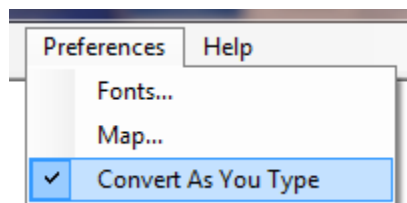


FIGURE 7 CONVERT AS YOU TYPE

Convert as you type mode

In this mode, text is converted as you type into the editor and the typed letters are visible on the status bar until a terminating key such as a space or delete character is entered. Double click this area if you need to capture this typed text into the editor's paste buffer.

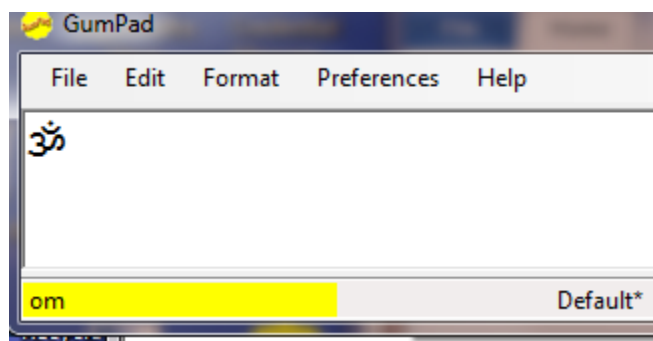


FIGURE 8 STATUS BAR SHOWING TYPED TEXT

The target language is visible on the right of the status bar. It can be changed by clicking on the little arrow that appears next to the name of the current language.

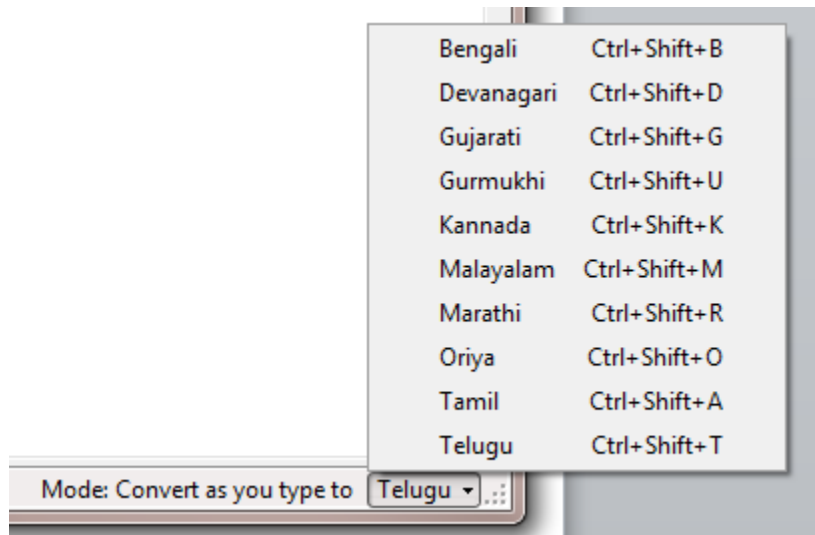


FIGURE 9 STATUS BAR SHOWING CONVERT AS YOU TYPE MENU

Convert after you type mode

In this mode, text appears in English as you type into the editor. After typing some text you can select all or parts of it and convert it from the Convert menu.

Recognized file extensions

GumPad recognizes the following file types. It will prompt you to specify an encoding if the file name extension does not match one of these types.

Extension	File Type
.gpd	Files in Rich Text format.
.txt	Plain text files in US-ASCII format.
.utx	Plain text files in Unicode format.
.itx	Plain text files encoded in ITRANS format.
.rtf	Rich Text Format files.

Add-In for Microsoft Word

GumPad is published as a standalone editor like Windows Notepad as well as an Add-In for Microsoft Word 2007 and Word 2010. The add-in appears as a *Transliteration* group on the *Home* tab of the Word Ribbon. To convert text, select it, click the *Convert* button and choose the language from the drop down list.

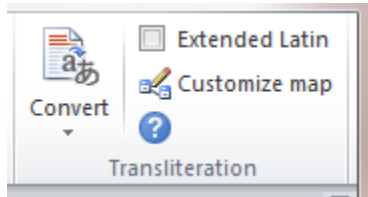


FIGURE 10 WORD 2010 TRANSLITERATION GROUP

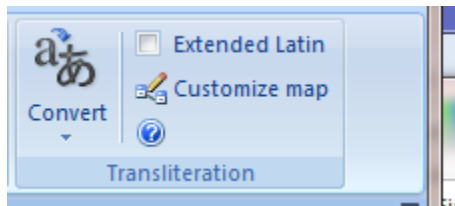


FIGURE 11 WORD 2007 TRANSLITERATION GROUP

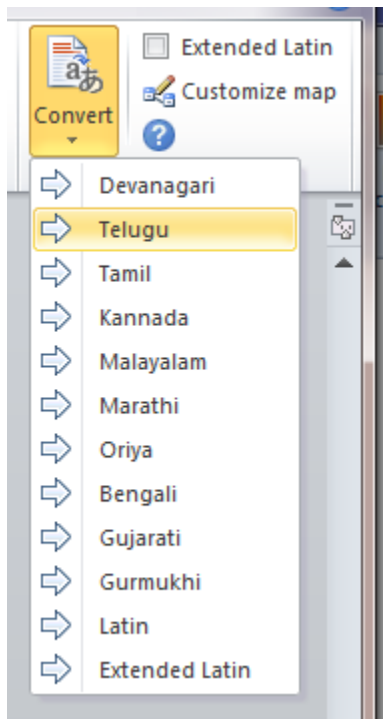


FIGURE 12 WORD 2010 CONVERT BUTTON

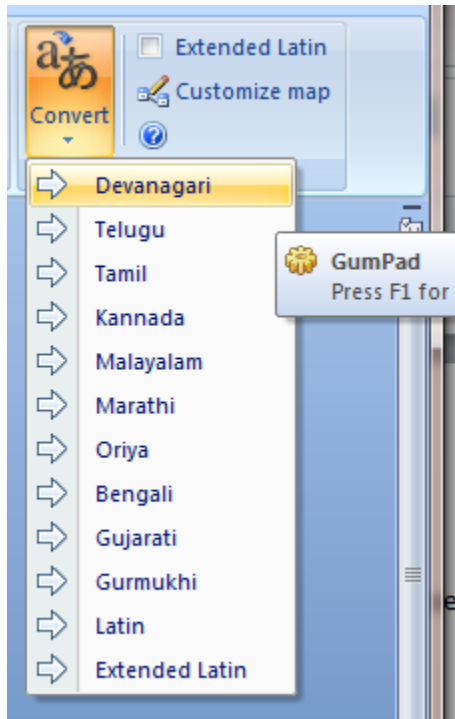


FIGURE 13 WORD 2007 CONVERT BUTTON

Use the *Extended Latin* checkbox to convert from Extended Latin and the *Customize Map* button to customize the conversion map.