

Unicode and IPA

Zev Handel, AL&L

The old days (i.e. the 1990s)

Fonts contained 256 glyphs, with code points numbered 0 to 255. Each code point was associated with a particular keyboard input. For example, <shift-a> would enter the glyph associated with code point 65. In most fonts, code point 65 was the glyph **A**. An IPA font would have a different set of symbols associated with the 256 code points. For example, in a particular IPA font code point 65 might be the glyph **ɶ**, and you would enter it by typing <shift-a>. The disadvantages of this method are obvious; if you send a file to someone who doesn't have your font, what you intended to be **ɶ** will be displayed as **A**.

Nowadays (Unicode)

Unicode is an encoding with 65,536 code points. This is enough that each distinct glyph can be assigned to a different code number. For example, **A** is number 65 and **ɶ** is number 336. Few fonts include all Unicode glyphs. A font can be any size. For example, you might have a Unicode IPA font with just 80 glyphs, but each will be associated with its Unicode code point. While not every font will contain the glyph **ɶ**, any font that does will associate it with 336. Sharing phonetic transcriptions is no longer font-dependent.

What you need to use Unicode IPA fonts

- 1) An operating system that understands Unicode, such as Mac OS X or Windows XP.
- 2) An application that is "Unicode-savvy", i.e. knows how to work with Unicode fonts, such as Microsoft Word.
- 3) A Unicode font that includes the glyphs of the IPA. Mac OS X and Windows XP both come with one complete Unicode font that contains (almost) all the Unicode glyphs. The Mac font is "Lucida Grande" and the Windows font is "Arial Unicode MS" or "Lucida Sans Unicode".
- 4) An input system that associates mouse-clicks or keyboard taps with the IPA symbols' Unicode codes.

Good Unicode IPA fonts

While the pre-installed Unicode fonts both contain the IPA glyphs, they are not aesthetically appealing. For example, this is what Lucida Grande looks like: **abcdɶɹðɲɥ**. The good thing about them is that everyone has them, so if someone to whom you send a file doesn't have a specialized IPA font, the IPA symbols will automatically show up in one of these fonts. Ugly, but not inaccurate.

The three best free Unicode fonts containing IPA symbols are **Doulos SIL**, **Charis SIL**¹, and **Gentium**, all available from SIL. They will function on either Windows or Mac. Doulos SIL blends well with Times and Times New Roman fonts.² Charis SIL is probably better for camera-ready copy, since it includes designed bold and italic faces. The Gentium font comes in two flavors; the "Gentium Alt" version is supposed to have "flatter diacritics" so that diacritics can be stacked attractively. All three fonts also contain all the ordinary Roman letters, so you can compose entire documents exclusively in one of these fonts.

Doulos SIL:	<i>abcdɶɹðɲɥ</i>	<i>abcdɶɹðɲɥ</i>	abcdɶɹðɲɥ
Charis SIL:	<i>abcdɶɹðɲɥ</i>	<i>abcdɶɹðɲɥ</i>	abcdɶɹðɲɥ
Gentium:	<i>abcdɶɹðɲɥ</i>	<i>abcdɶɹðɲɥ</i>	abcdɶɹðɲɥ

These fonts can be downloaded at:

¹ Note these Doulos and Charis fonts are different from the older, non-Unicode SIL IPA fonts with similar names.

² Unfortunately, the line spacing of the Doulos font is problematic. This can be corrected by selecting an "exact" line spacing in the "Format Paragraph" panel in Word.

http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&item_id=CharisSILfont
http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&item_id=DoulosSILfont
http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&item_id=Gentium

Inputting IPA symbols: Macintosh

The biggest challenge to using Unicode IPA fonts is getting the symbols into your documents. This is because there is no fixed association of keyboard strokes with the code points associated with the IPA glyphs.

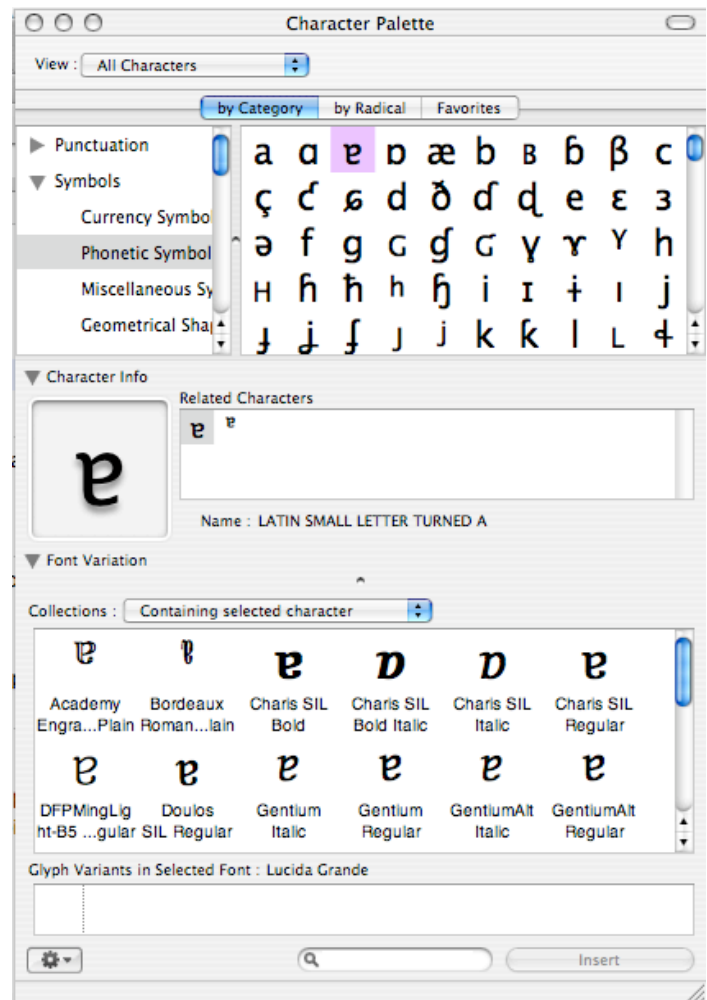
On a Mac, there are two keyboard utilities that you should install. The first, the **Character Palette**, is built into the operating system. To install it, open “System Preferences”, select “International”, choose the “Input Menu” tab, and check “Character Palette”. (You can also do this by choosing “Open International...” under the flag menu in the upper right of the screen.) You will now have the powerful “Show Character Palette” option in your flag menu.

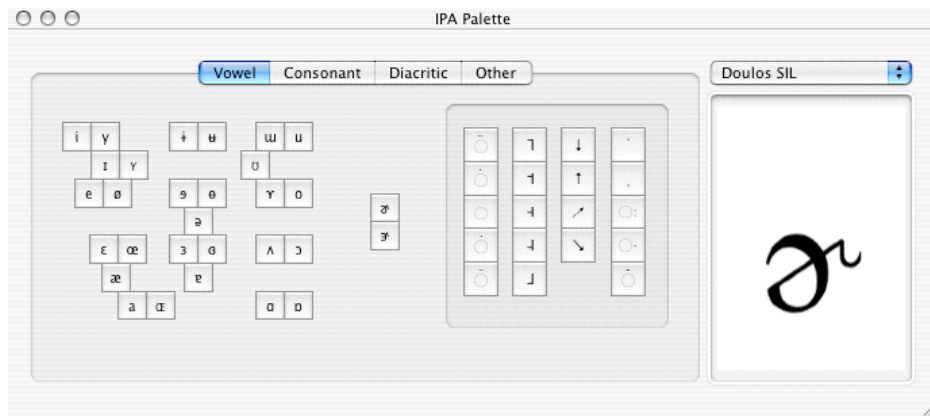
The Character Palette contains the complete set of Unicode glyphs, and you can view them by code number or by “category”. Categories are logical groupings such as “Hebrew alphabet”, “Punctuation”, “Currency Symbols”, “Braille”, etc. The Character Palette will show you which of your available fonts can display a particular glyph, and point you to related glyphs. You can add frequently used glyphs to a “Favorites” section to make them easier to find later, and enter glyphs directly into your document from the palette. (Note: The Character Palette looks somewhat different in different versions of Mac OS X. The image here is from OS 10.4.)

In OS X.3, the IPA symbols are distributed over four different “blocks”: “Latin-1 Supplement”, “Latin Extended-A”, “Latin Extended-B”, and “IPA Extensions”. In OS X.4, they are consolidated under the category “Phonetic Symbols”.

The second keyboard utility, **IPA Palette**, is available free for download at <http://www.blugs.com/IPA/>. Once installed, you activate it the same way that you activate the Character Palette, by checking it in the “Input Menu” tab of the “International” pane. This utility lays out all the IPA symbols in the format familiar to linguists. Mousing over one will show you what it looks like in any font you select. Clicking on a symbol inputs it into your document. (See image on next page.)

For an IPA keyboard that works with Mac, go to <http://stedt.berkeley.edu/stedtfon/stedtfon.html> and find the link to download the **STEDTKeys** keyboard. Once installed you select it from the flag menu. You will need to consult the documentation to know which keys are mapped to which IPA characters.





If you prefer to enter Unicode characters by typing the hexadecimal code, you can install and use the **Unicode Hex Input** keyboard that comes with Mac OS X, which is also found in the International panel. To enter characters when this keyboard is selected, hold down the option key while typing the four-digit hexadecimal code number. For example, opt-0041 will give A, and opt-025a will give æ.

Inputting IPA symbols: Windows

You can download from SIL a keyboard utility that associates IPA characters with particular keystrokes at http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&id=IPAhome (click “IPA Unicode Keyboards”), or try the free Unicode phonetic keyboard from <http://www.phon.ucl.ac.uk/resource/phonetics/>. You can also enter Unicode characters using the Unicode code number by typing the four-digit number of the numeric keypad while simultaneously holding down the ALT and X keys. (Note: Unicode codes are often given in hexadecimal format. These must first be converted to decimal format before they can be entered this way!)

Or, if you primarily use Microsoft Word, you can choose “Symbol...” from the “Insert” menu, and scroll down through the Unicode blocks to locate the IPA symbols you frequently use. Once you locate one, you can assign a keystroke to it, or make an “auto-correct shortcut”. However, these keystrokes will not apply to other applications.

This page has some useful tips for entering IPA symbols in Word on Windows and explains how to make “auto-correct shortcuts”: <http://www.phon.ucl.ac.uk/home/wells/phoneticsymbols.htm>