

# What is this magic setting that synthesizes Unicode from non-Unicode?

[devblogs.microsoft.com/oldnewthing/20090115-00](http://devblogs.microsoft.com/oldnewthing/20090115-00)

January 15, 2009



Raymond Chen

Commenter dan g. wonders how Windows can treat non-Unicode applications as Unicode via the Regional and Language Options control panel, specifically the part that lets you choose the *Language for non-Unicode programs*. “Having always believed that the only way to display, say, Chinese characters correctly was to compile with `_UNICODE`, this facility seems all the more remarkable.”

This setting is really not as magical as it appears. (After all, we had Chinese versions of 16-bit Windows that displayed Chinese characters just fine, and they certainly didn’t use Unicode since Unicode hadn’t been invented yet.) Michael Kaplan went through this and many other settings in the Regional and Language Options control panel, and from the chart at the top of the page, you see what Windows XP calls the *Language for Non-Unicode Programs* used to go by the name *Default System Locale*. The old name does a better job of describing what it actually does but does a worse job of describing what it’s used *for*.

In Win32, three character encodings have special status. Unicode (more precisely, UTF-16) of course is what Windows uses internally. There are also two 8-bit code pages: `CP_ACP`, the so-called *ANSI code page* (even though it isn’t actually ANSI), and the `CP_OEM` code page, the so-called *OEM code page* (even though it isn’t provided by the OEM).

When a non-Unicode program calls a function like `TextOutA` to display a string represented in the ANSI code page, the string is converted to Unicode via the `CP_ACP` code page. The *Language for non-Unicode programs* setting controls what code page `CP_ACP` corresponds to. On U.S. systems, it’s typically code page 1252, but you can change it via that control panel. And that’s where it becomes possible to display Chinese characters without using Unicode.

For example, code page 950 is a double-byte code page commonly seen in countries that use traditional Chinese characters. It can represent the English alphabet of A-Z, and through the use of double-byte characters can also represent a wide array of traditional Chinese characters, such as this block of characters which are represented by byte sequences of the

form B3\_40 through B3\_FE. If the ANSI code page is code page 950 and you pass data formatted for that code page to, say, the `TextOutA` function, the corresponding Chinese characters will display, even though the program itself doesn't use Unicode explicitly.

That's why it's called the *Language for non-Unicode programs*. It specifies which character set non-Unicode data should be interpreted as.

Raymond Chen

**Follow**

