Kazuomi Kuniyoshi
Japanese formatting rules for XeTeX

Pravidla sazby japonštiny v XƎTEXu

Kazuomi Kuniyoshi

Abstrakt

Tato krátká zpráva komentuje důvod vzniku balíčku genzi, který nastavuje a upravuje pravidla sazby japonštiny v XƎTEXu. Komentáře, balíček i několik ukázek si lze stáhnout z webové stránky autora, viz http://kuniyoshi.fastmail.fm/xetex/. Balíček zatím není součástí CTAN.ORG.

Klíčová slova: ČJK, XƎTEX, japonština, balíček genzi.

In principle, XƎTEX lets you mix as many writing systems as you wish in a document. With an appropriate OpenType font being installed in the operating system, you can freely use all the letters which the font contains. Moreover, you can put letters from different fonts onto a document at the same time. This is a marvellous feature of XƎTEX.

I am a multilingual writer, or to be more precise, multi-*script*-al writer, who loves the power of TEX. The writing systems I often mix are Japanese and Latin. It is no big deal to do so in today's word processors, but fine-tuning the balance of these letters, let alone producing an overall beautiful layout, is not an easy task in such an environment. So the advent of XƎTEX was an exciting news for me. Yet, on trying it for the first time, I immediately faced a problem: XƎTEX does not know the formatting rules for Japanese writing well.

Some websites give an impression that XƎTEX can produce good Japanese documents only by touching `\XeTeXlinebreaklocale` and `\XeTeXlinebreakskip`. This is simply false; everyone who knows how to write in Japanese should be able to see quickly that this is like a paradise (mis)represented in a summer travel brochure. Indeed, XƎTEX can print Japanese letters; however it ignores the formatting rules to the extent that it cannot be used in serious publications. It seems to me that there is some confusion over printing letters and formatting them among developers and users.

In the world of computing, there is a convenient word 'CJK'. This word is convenient as long as letters are concerned, but it turns out to be inconvenient as soon as document formatting becomes the topic of discussion. Perhaps it is right to say that 'C', 'J' and 'K' are similar enough to be treated together for engineering purposes. Chinese and Japanese share a lot of letters despite subtle differences, and it is unnecessary to come up with different engineering methods to deal with them. I totally agree with this.

XƎTEX is, however, not only a tool to print letters onto PDF files (and sometimes to sheets of paper), but also a powerful formatting tool. In any event, that is why people are drawn into TEX most of the time. So it is important to separate printing letters and formatting them; for printing letters is an engineering issue, but formatting them is a cultural issue. And, 'C', 'J' and 'K' do differ greatly when it comes to their formatting rules.

The current state of XƎTEX is almost perfect on the printing side. There is wonderful freedom in picking up any of 15,444 letters from an Adobe-Japan1-4 font in XƎTEX. But whenever I write a document in Japanese or (Classical) Chinese, XƎTEX does not automatically format text correctly. This is an area where more work should be done, and it is interestingly an interdisciplinary one spanning software engineering, linguistics and literature.

The genzi package I wrote is a preliminary attempt to bridge this gap.

```
http://kuniyoshi.fastmail.fm/xetex/
```

Those who use other writing systems in XƎTEX, especially very 'exotic' ones, may also be feeling that cultural differences should be considered more. As TEX, which was exclusively American in the beginning, has gone out of the Euro-American world and stepped into the dangerous 'polyscript' terrain of East and Southeast Asia, more humanism should melt into its engineering than ever before.

Summary: Japanese formatting rules for XƎTEX

This is a two-page report informing about the reasons and the existence of the genzi package which sets Japanese formatting rules for XƎTEX. The package, samples and more comments can be viewed and downloaded from the author's website, see http://kuniyoshi.fastmail.fm/xetex/.

Keywords: CJK, XƎTEX, Japanese, genzi package.

Kazuomi Kuniyoshi, Brussels, Belgium
kuniyoshi@fastmail.fm
CₛTUG c/o FEL ČVUT, Technická 2
Prague, CZ-166 27, Czech Republic