

X_ƎLaTeXで日本語する件について

ZR

2010 年 某月 某日

1 X_ƎTeX の紹介

1.1 X_ƎTeX とは

X_ƎTeX は Jonathan Kew 氏による TeX の拡張で符号空間を Unicode 全体 (BMP 以外も含む) に拡大したものである。既存の TeX の Unicode 拡張としては Omega (Ω) があるが、X_ƎTeX はこれとは別の拡張となっている。(なお、Omega の後継にあたる Unicode 拡張 TeX が LuaTeX である。)

X_ƎTeX のもう一つの大きな特徴は、現在の標準的なフォント技術を内部に取り込んでいることである。元々の TeX が「特定の技術基盤に依存しない」という思想で設計されていることはよく知られているが、これはフォント関係の技術についても当てはまり、フォントの扱いは TeX 本体 (すなわち TeX ソースを DVI ファイルに変換する) ではなく DVI ウェア (DVI ファイルを扱うソフトウェア) が行う。このため、TeX の機能自体は 30 年前からほとんど変わっていないにも関わらず、dvipdfmx と組み合わせて OpenType フォントを埋め込んだ PDF 文書を作成することが可能になっている。しかしその代償として、例えば新しいフォントを使えるようにするために一定の作業が必要になるなどの短所を持ち合わせている。X_ƎTeX はこの点に関して「非依存性」を捨てて、システムで使用可能なフォントを直接扱うことで、何も設定しなくても好きなフォントを使うことが可能になっている。また、従来の TeX では難しい、フォントのもつ情報を利用した高度なりガチャ・アクセント付加・位置異形の処理も自由に使うことが可能になっている。

1.2 X_ƎTeX の出力例

X_ƎTeX のもつ能力の一端を示す例をあげる。これは X_ƎLaTeX (X_ƎTeX 上で動く LaTeX) の文書である。

```
% このファイルの文字コードは UTF-8
%documentclass[a4paper]{article}
%usepackage{xltextra} % これは後で解説
%newfontfamily%fchr{Charis SIL} % %fchr でフォント “Charis SIL” に切替
%newfontfamily%fipm{IPA明朝} % %fipm でフォント “IPA明朝” に切替
% %faru でデーヴァナーガリー文字用の設定を施した “Arial Unicode MS” に切替
%newfontfamily%faru[Script=Devanagari]{Arial Unicode MS}
% %showUC は何の変哲のない LaTeX のマクロ
%newcommand*%showUC[1]{% %texttt{%footnotesize U+#1}}
%begin{document}
%begin{itemize}
```

```

%item {\fipm 土\symbol{"571F}}\showUC{571F} /
      {\fipm 土\symbol{"5721}}\showUC{5721} /
      {\fipm 土\symbol{"2123D}}\showUC{2123D}
%item {\fchr j}\showUC{006A} + {\fchr\symbol{"0302}}\showUC{0302}
      = {\fchr j\symbol{"0302}}
%item {\fipm か}\showUC{304B} + {\fipm\symbol{"309A}}\showUC{309A}
      = {\fipm か\symbol{"309A}}
%item {\faru \symbol{"915}}\showUC{0915} + {\faru\symbol{"93F}}\showUC{093F}
      = {\faru \symbol{"915}\symbol{"93F}}
%end{itemize}

%end{document}

```

上の文書を xelatex コマンドで組版すると以下の出力を含む PDF 文書が得られる。

- 土土 U+571F / 土土 U+5721 / 土土 U+2123D
- j U+006A + [^] U+0302 = \hat{j}
- か U+304B + [°] U+309A = か°
- क U+0915 + ि U+093F = कि

上の文書において、`\symbol{"xx}` は符号位置が xx (16 進) である文字を出力する L^AT_EX コマンドである。X_ƎL^AT_EX ではこれが Unicode 符号空間で使えることが出力の 1 行目から判る。(従来の L^AT_EX では 8 ビットの値のみ。) 勿論、UTF-8 で直接書いた文字も出力されている。2 行目以下では Unicode 結合文字を用いた文字合成の例を示している。j`\symbol{"0302}` は j の直後に U+0302 の文字 (結合曲アクセント) を書いたのと同じで、これは Unicode の規則で「 \hat{j} (曲アクセント付 j)」を表す。2 行目を見ると正しく合成が行われていることが判る。