

1 XeTeX とは

The XeTeX typesetting system は UTF-8 ベースの TeX。「ズイーTeX」と読む。Mac OS X 上で開発され、システムが認識する OpenType フォントをそのまま使える。Linux 版もある。W32 版は Linux 版を移植したものである。Linux と W32 では、使えるフォントは OS が認識しているものと言うより、FontConfig が認識しているものと言ったほうが正確である。

(中略)

デフォルトではタイプセット時に、パイプによって出力がドライバに送られ、最終出力は PDF となる。オプション “--no-pdf” によってこれを 2 段階に分けることもできる。この場合は .xdv ファイル (.dvi ファイルを拡張したもの) を出力する。xdv ファイルをドライバによって PDF に変換する。xdv2pdf は Mac 専用のドライバであって、他では利用できない。Linux と W32 では、汎用の xdvipdfmx というドライバを用いる。これは dvipdfmx を XeTeX 用に拡張したものである。xdv2pdf は Mac OS X を有効利用するドライバだが、仮想フォントを読むことができないので、通常の tfm, vf 形式の TeX, LaTeX 用フォントも利用する場合には Mac OS X であっても xdvipdfmx を使うことになる。また、上記の “Uninstalled fonts” を使う機能も、ドライバに xdvipdfmx を使う場合だけ有効である。直接 PDF を出力する場合には

```
xelatex --output-driver='xdvipdfmx -q -E' foo
```

のようにすると、デフォルトのドライバを変更して xdvipdfmx を使うようにできる。なお Mac OS X 用の XeTeX には最新の 0.995 以降のものにしか xdvipdfmx は含まれていない。svn revision 506 以降では、Mac OS X 用のデフォルトのドライバも xdvipdfmx になった。パッケージに同梱されている graphics.cfg と color.cfg は xdv2pdf 用の設定がなされている。従って Linux と W32 では、graphics パッケージや color パッケージなどを使用するとき、必ず [dvipdfm] オプションを付加しておくのを忘れてはいけない。ただし、Jonathan Kew さんにより、Mac と Linux/W32 とでインターフェイスを統一する仕事が行われており、最新版の xdvipdfmx と XeTeX を W32 版でテストしてみると、デフォルトのままで xetex.def を使えるようになっている。

しおり等の PDF 文字列に関しては、オリジナルの dvipdfmx では適切な CMap を指定することで Big Endian UTF-16 with BOM に変換するようになっていた。xdvipdfmx でもこの機能はそのまま使えるが、UTF-8 で入力する XeTeX においては単に UTF-8 を Big Endian UTF-16 with BOM に変換するだけでよいので、より簡単な実装が可能である。実際にこの機能が実装されていて、特別に何もしなくても日本語しおり等を作ることができる。

2 改行アルゴリズムの比較

そもそも TeX で改行が可能な位置は、(数式と kern を無視して) 単純化すると、glue、penalty、discretionary break です。ハイフンは日本語に関係しないので、無視します。文章中のどのような箇所に glue と penalty が現れるかというと、

1. ユーザーが直に書き込んだ場所、
2. マクロの展開の結果書き込まれる場所、
3. それ以外、

がありますが、ここでは、3. が問題となります。

- 古典的な 8 bit 版の TeX では、空白文字の箇所に TFM から読みとられた量の glue が自動的に挿入されます。
- pTeX では、更に、和字と和字の間に TFM (JFM) から読みとられた量の glue、または、\kanjiskip が、和字とそれ以外の間に \xkanjiskip が自動的に挿入されます。また、禁則文字の前後には penalty が自動的に挿入されます。このグルーの挿入はマクロ展開後に行われます。更に詳しいことや、カスタマイズの方法については、『pTeX での JIS X4051 の実装』を御覧下さい。
- XeTeX では、空白文字の箇所に加え、\XeTeXlinebreaklocale が有効なロケールになっていれば、ユニコードの改行規則に基づいた改行可能箇所を内蔵の ICU ライブラリが探し出し、\XeTeXlinebreakpenalty と \XeTeXlinebreakskip が自動的に挿入されます。例えば、

```
\XeTeXlinebreaklocale "th"
\xeTeXlinebreakpenalty 0
```

とすれば、タイ語で書かれたテキストを ICU ライブラリに内蔵された辞書を用いて、単語に分割し、その区切り目に、0 のペナルティを挿入します。一方、上にも紹介されている

```
\XeTeXlinebreaklocale "ja"  
\XeTeXlinebreakskip=0em plus 0.1em minus 0.01em
```

は、和字と和字の間や和字とそれ以外の間に、少量の伸び縮みを許した幅 0 の空白 (glue) を挿入することを指示します。この際、行頭禁則文字の直前と、行末禁則文字の直後は、ユニコードの改行規則では改行可能であると認識されないので、このような場所には何も挿入されません。

このようにして、各言語毎の規則で設定された改行可能箇所から、その段落の最も適切な改行箇所を見つけ出すのは、 $\mathrm{T}_{\mathrm{E}}\mathrm{X}$ の役割です。0.997 以降では、更に、`\XeTeXinterchartokenstate` が正の値になっていれば、`\XeTeXcharclass` の概念が有効になり、`\XeTeXcharclass` が A の文字と `\XeTeXcharclass` が B の文字との間にはトークンレジスタ

```
\XeTeXinterchartoks A B
```

が自動的に挿入されます。このトークンレジスタに改行を可能にする glue や penalty に展開される内容が入っていれば、ここの改行可能箇所になります。このグルーの挿入はマクロ展開前に行われます。`\XeTeXcharclass` は 0 ~ 255 の任意の整数ですが、255 だけは特別な意味を持ち、文字列の境界を表します。

例えば、行末、行頭、インラインの数式、hbox などです。FMT ファイルで、`\XeTeXcharclass` の値は

- 分かち書きされる通常の文字は 0
- 漢字と仮名は 1
- 開き括弧は 2
- 閉じ括弧と句読点は 3
- 結合文字は 256 (無効な値、`\XeTeXinterchartoks` の挿入を不可能にするため。)

に初期化されます。これらの文字の分類は、Line Breaking Properties に基づいて行われますが、0 以外の値が設定されるのは、East Asian Width の定義で、Full Width, Half Width, Wide のいずれかの属性を与えられた文字に限られます。更に、

```
\gdef\txHanGlue{\hskip0pt plus 0.1em\relax} % between ideographs  
\gdef\txHanSpace{\hskip0.2em plus 0.2em minus 0.1em\relax} % before/after runs of CJK  
  
\global\XeTeXinterchartoks 0 1 = {\txHanSpace}  
\global\XeTeXinterchartoks 1 0 = {\txHanSpace}  
\global\XeTeXinterchartoks 1 1 = {\txHanGlue}  
\global\XeTeXinterchartoks 1 2 = {\txHanGlue}  
\global\XeTeXinterchartoks 1 3 = {\nobreak\txHanGlue}  
\global\XeTeXinterchartoks 2 1 = {\nobreak\txHanGlue}  
\global\XeTeXinterchartoks 3 1 = {\txHanGlue}
```

等と、トークンレジスタも設定されていますが、これらは自由に再設定できます。 $\mathrm{pT}_{\mathrm{E}}\mathrm{X}$ に於ける `\inhibitglue` に相当する処理が必要ならば、`\relax` を使えばよいでしょう。(ただし、通常の文字とクラス 255 の文字との間に 0 ではないグルーが挿入されていない前提です。) 一例として、行末禁則文字の `\XeTeXcharclass` を 4 に設定するには、

```
\@for\Kinsoku:={`#`,`¥`,`$`,`&}\do{\XeTeXcharclass\Kinsoku=4}
```

とします。`\XeTeXinterchartoks 4 1` 等を適切に定めれば、これらの文字の直後での改行を抑制できます。`\XeTeXinterchartoks` にはフォントの切替え命令を設定することもできます。これにより、dvi ドライバではなく、 $\mathrm{T}_{\mathrm{E}}\mathrm{X}$ 側に於ける、バーチャルフォントが実現できます。なお、このメカニズムにより、文字間にトークンを挿入すると、内蔵の ICU ライブラリによる解析の対象は個別の文字になってしまうので、`\XeTeXlinebreaklocale` の設定は無意味になります。同様に、OpenType フォントの持つ、コンテキスト依存の文字置換や、リガチャーを使用する際にも、`\XeTeXinterchartokenstate=0` としなければなりません。