

外字とは、以下の意味で使われる言葉です。

- ・ 文字コード表にない文字。
- ・ 上の意味の外字を、文字コードの空き領域に独自に割り当てたもの。ユーザが設定する ユーザ定義外字と、ベンダが製品に組み込んでいる ベンダ定義外字とがある。

例えば、丸付き数字は JIS X 0208 にないので、この 文字コードにとっては、符号化できない外字となります。これが上記1番目の意味です。(JIS X 0213 や Unicode には 丸付き数字 はあるので、これらの 文字コード にとっては外字ではありません)

また、Windows や Mac OS といったソフトウェア製品では、Shift JIS の空き領域に独自に 丸付き数字 を割り当てたものがあります。これが上記2番目の意味の外字 (ベンダ定義外字) です。

Unicode では私用領域 (Private Use Area) が定められており、ここに外字を割り当てることができます。

外字の問題点とその解決方法

ユーザやベンダが独自に割り当てた外字は他の製品や環境とは互換性がありません。いわゆる文字化けの元となります。 JIS X 0208 の空き領域を使用した ベンダ定義外字 は昔からトラブルの元となっています。

丸付き数字 や ローマ数字 等を取り込んだ 文字コード標準、つまり JIS X 0213 や Unicode を使うことで、外字のトラブルを無くすことができます。

例えば、UTF-8 と SJIS の間でコード変換をするときには、JIS X 0208 ベースの Shift JIS でなく、拡張規格の JIS X 0213 ベースの Shift JIS-2004 を指定すれば、丸付き数字 や ローマ数字 はもちろん、専門用語や地名などに使われる第3第4水準漢字や、アクセント付きの ラテン文字、英和辞典に用いられる発音記号、著作権表示記号など各種記号類も文字化けせずに使用できます。 iconv コマンドでコード変換するときはこうします：

```
iconv -f UTF-8 -t SHIFT_JISX0213 < utf8.txt > sjis.txt
```

関連項目

- ・ 国内実装互換