

Shift JIS (シフト JIS、略して SJIS とも) は、JIS X 0208 を JIS X 0201 と一緒に使うための符号化方式です。JIS X 0208 の附属書 1 で定義されています。上位互換の拡張規格として Shift JIS-2004 があります。

## 符号の構造

JIS X 0201 の 8 ビットコードは、0x7F 以下 (GL 領域) にラテン文字集合 (ASCII と類似、2 文字だけ違う) を、0xA0 以上 (GR 領域) に片仮名集合の、ともに 1 バイトコードを配置しています。しかし片仮名集合は 0xE0 以降が空いているので、そこと 0x80-0x9F の制御文字領域をも利用して、JIS X 0208 の文字を詰め込んでしまったものが Shift JIS です。

元々の JIS X 0208 の格好では JIS X 0201 の隙間に詰め込むことができないので、計算によって変形しています。

JIS X 0208 の区点番号から Shift JIS の第 1・第 2 バイトを計算する方法は、下記の Shift JIS-2004 の計算式を参照してください。面番号が 2 の場合を無視すると Shift JIS と同じです。

- ・ JIS X 0213 の代表的な符号化方式 #Shift JIS-2004

## 使用上の注意事項

注意点として、2 バイト文字の第 2 バイトの範囲が、1 バイト文字である JIS X 0201 ラテン文字と重なる (つまり、0x7F 以下の範囲を含む) ことが挙げられます。

特に、メタ文字としてしばしば用いられる 0x5C (円記号、ASCII ではバックスラッシュ) もシフト JIS の第 2 バイトになり得ることから、以前より問題を引き起こしてきました。

## 他のコードとの関係

### Shift\_JIS-2004

Shift JIS は Shift JIS-2004 のサブセットとみなせません。別の言い方をすれば、Shift JIS-2004 をサポートすれば、Shift JIS をサポートしたことにもなります。

### Unicode

Shift JIS と Unicode との機械可読な変換表は、当サイトの「JIS X 0213 のコード対応表」にあります。このページからダウンロードできる「Shift JIS-2004 と Unicode の対応表」のファイルがそのまま使えます。[2000] 及び [2004] と記された行を読み飛ばせば、JIS X 0208 ベースの Shift JIS にある文字だけに限定されます。

### ベンダ定義外字 (CP932, Windows-31J, MS932 等)

Shift JIS には各社のベンダ定義外字を空き領域に追加した亜種がありますが (例えば Windows の CP932 ないし Windows-31J)、規格違反の実装であり Shift JIS-2004 との互換性がなく (一部の丸付

き数字等の非漢字除く)、また JIS X 0208 の欠点を補うには量・質ともにあまりにも不十分であり、二重符号化(ひどい場合には三重符号化)の問題さらには Unicode との変換がおかしいという問題を持つものもあるため、使うべきではありません。

## 問題点とその解決法

Shift JIS にはいくつかの問題があります。

- ・ JIS X 0208 をベースにしているため、文字が足りない
- ・ 空き領域にベンダが独自に文字を割り当てたものがあり、相互運用性の問題を生じている(ベンダ定義外字、いわゆる機種依存文字)
- ・ ベンダの実装によっては、Unicode との変換において、いくつかの記号が JIS の定義と異なる間違った文字に変換するものがある(波ダッシュ問題)

これらの問題は、Shift JIS の上位互換の符号化方式である Shift JIS-2004 を使うことで解決できます。

コード変換の際に、例えば iconv ではこう指定します。

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

この "SHIFT\_JIS" を "SHIFT\_JISX0213" に置き換えれば、Shift JIS-2004 が使われ、上記の問題が解決します。

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

反対方向、つまり UTF-8 から SJIS に変換するには、-f と -t を逆にします：

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

なお、SHIFT\_JISX0213 を使うと、Windows のベンダ定義外字の丸つき数字やローマ数字等の 13 区の記号類も救済することができます。

## 関連項目

- ・ Shift JIS-2004 - 拡張版
- ・ JIS X 0208 - 元になっている規格
- ・ EUC-JP - JIS X 0208 の符号化方式の一種
- ・ ISO-2022-JP - JIS X 0208 の符号化方式の一種