

漢字コードについて

理学部情報科学科 八重樫 裕 幸

slyae@si.hirosaki-u.ac.jp

パソコンを利用して通信を行った方なら必ず経験しているだろう、文字化けについてとWSで使用されている漢字コードのいろいろについて簡単に説明します。

漢字コードにはシフトJIS (SJIS)、JIS、EUC等が在ります、JISでは漢字IN、漢字OUTのコードが入ってきます。この漢字IN、漢字OUTのコードにも色々種類が在ります。

シフトJISは主にパソコンで使用されている漢字コードです。EUCはWS等で使用されている漢字コードです。また電子メール及びニュースシステムの文字コードはJISが標準になっています。シフトJIS・EUCは8 bit系、JISは7 bit系のコードです。

通信ソフトで文字化けする場合の主な理由は漢字コードの不一致です。

WS (zws0) ではnkfというテキストファイルの文字コード変換してくれるプログラムがあります。WSからパソコンにファイル転送する場合はあらかじめ転送ファイルのコードをnkfでsjisに変換しておくことを勧めます。

```
nkf -s 文件名 > 出力file名
```

nkfプログラムでは次の漢字IN・漢字OUTをサポートしています。

漢字IN ---- “ESC \$ B”、 “ESC \$ @”

漢字OUT ---- “ESC (J”、 “ESC (B”、 “ESC (H”

例) 漢字IN--@ 漢字OUT--JのJISfileをeucコードに変換する場合。

```
nkf -e -@J 文件名 > 出力file名
```

通信ソフトを利用する場合は、パソコンーWSはEUC、パソコンーパソコンはSJISを使用することが無難な選択です。

また半角カタカナはWSで使用できないと考えてください。

ネットワークソフト (mail、news等) では半角カタカナを利用しない事になっています。必要に迫られた場合は、8ビットオプションが利用できるeucコード変換プログラムを利用すれば可能になるでしょう。(SUNのWSの場合、sjtoeuc・jistoec等)

トラブルを避けるためには半角カタカナは利用しない方がいいです。