

## 補足 : Windows コードページと標準的な文字エンコード方式の違い

Microsoft は、独自で文字エンコード方式を定義開発して自社製品に実装しており、Windows では各地域の文字セットを「コードページ」Codepage として番号で表している。コードページの収録する文字が、対応する文字エンコード方式の標準規格と異なることがあるので注意が必要である。

## この授業で扱う言語・地域で標準的に使われるエンコード方式とコードページ

言語(地域)	標準的な文字エンコードの名称	Windows での名称	Windows コードページ Codepage
日本語	Shift JIS	日本語 (シフト JIS)	932
西ヨーロッパ言語	ISO-8859-1	西ヨーロッパ言語(ISO) 西ヨーロッパ言語(Windows)	1252
中国語 (中国)	GB2312	簡体字中国語 (GB2312)	936
中国語 (台湾)	BIG5	繁体字中国語 (Big5)	950
韓国語	EUC-KR	韓国語 (EUC) 韓国語	949
タイ語	TIS-620	タイ語 (Windows)	874

(第 6 回資料参照)

詳しいコードページの情報には、<http://www.microsoft.com/globaldev/reference/WinCP.asp> に詳しいので、参照するとよい。Windows の主なコードページは以下のとおり。

- 1 バイト言語のコードページ
 

1250 (中央ヨーロッパ言語, Central Europe)	1256 (アラビア語, Arabic)
1251 (キリル言語, Cyrillic)	1257 (バルト言語, Baltic)
1252 (西ヨーロッパ言語, Latin 1 *)	1258 (ベトナム語, Viet Nam)
1253 (ギリシャ語, Greek)	874 (タイ語, Thai)
1254 (トルコ語, Turkish)	
1255 (ヘブライ語, Hebrew)	

\* Latin 1 は ISO-8859-1, ないし Windows コードページ 1252 の通称。

- 2 バイト言語のコードページ

932 (日本語シフト JIS, Japanese Shift-JIS)  
 936 (簡体字中国語, Simplified Chinese GBK)  
 949 (韓国語, Korean)  
 950 (繁体字中国語, Traditional Chinese Big5)

外国語テキスト文書を正しく使うためには、標準的な文字エンコード方式とともに、Windows 独自のエンコード方式に注意する必要がある。次ページの表は、コードページ 1252 (Latin 1), 西ヨーロッパ言語のコードページである。ISO-8859-1 のコード表と比べ、違いを調べよう。

コードページ 1252 (西ヨーロッパ言語) では、ユーロ記号や引用符のバリエーション、さらには若干のアクセント記号つきアルファベットなど、標準的な ISO-8859-1 のコードでは表現できない文字が使えることに注意しよう。これらは、Windows 上でのみ有効な文字コードで、Mac やその他の OS では文字化けしてしまう。逆にいうと、1987 年に制定されたエンコード方式 ISO-8859-1 では、ユーロ記号は使えない。

