

概要

CJK 統合漢字とは、ISO/IEC 10646 ならびに Unicode において、東アジアの漢字文化圏各国の漢字コード規格にある漢字を統合して作成された漢字集合です。CJK という呼び方は China, Japan, Korea の頭文字から来ています。

元々、20,902 文字からなる CJK 統合漢字が Unicode/10646 のために作成されました。その後、CJK 統合漢字拡張 A が BMP 内に追加され、ついで、面 02 に CJK 統合漢字拡張 B が追加されました。現在、拡張 H まであります。拡張 B から拡張 F までは面 02 にあり (つまり符号位置 U+2xxxx)、拡張 G, H は面 03 (U+3xxxx) に配置されています。

CJK 間の字体包摂

各国の漢字コード規格には、同じ漢字であっても、文字デザインの違い、あるいは字体に及ぶ違いが見受けられます。CJK 統合漢字では、形の違いの大きいものは区別せず包摂して単一の符号位置を与えています。例えば、「骨」という字の中のカギは中国簡体字の文字コード規格 GB 2312 の例示字形では左右逆についていますが、この差は CJK 統合漢字は区別せず同じ符号位置 U+9AA8 で表します。

原規格分離

ただし、最初に CJK 統合漢字を作ったときに元になった原規格で分離されている字体は、CJK 統合漢字でも包摂せずに分離されています。例えば、「説」の右上の点 2 つが下向きに開いた字体は、JIS X 0208 では包摂されていますが、台湾の規格が異なる符号位置を与えていたために CJK 統合漢字でも分離されています。これを原規格分離規則 (source separation rule) といいます。

互換漢字

CJK 漢字統合の包摂対象となる字体であるが他の文字コードとの往復変換 (round-trip conversion) のために必要なものは、CJK 互換漢字という形で対応されていることがあります。韓国規格 KS X 1001 で重複している漢字や、JIS X 0213 で人名用漢字の関係で追加された一部の漢字がこれに当たります。

JIS X 0213 との関係

JIS X 0213 の漢字は、CJK 統合漢字、CJK 統合漢字拡張 A、CJK 統合漢字拡張 B に入っています。

特に、BMP 外、面 02 の拡張 B にもあるため、JIS X 0213 の文字全てを正しく扱うには、UTF-16 ではサロゲート・ペアの、UTF-8 では 4 バイトのコードの処理が正しく出来る必要があります。

JIS X 0213 の漢字の一部、「人名許容・康熙別掲」で包摂規準の適用除外となる符号位置は CJK 統合漢字でなく CJK 互換漢字として Unicode に取り込まれているものがあります。例えば「福」のへんが「ネ」でなく「示」の字体 (面区点 1-89-33) がこれに当たります。

関連項目

- CJK 互換漢字
- ISO/IEC 10646
- Unicode