

文字種の取扱い基準 (2012年1月以降)

『JAPAN/MARC MARC21 フォーマット』に使用する文字種の取扱い基準

この文書では、『JAPAN/MARC MARC21 フォーマット』での文字の取扱いについて規定する。

なお、『JAPAN/MARC 2009 フォーマット』で使用する文字の取扱いは「[文字種の取扱い基準 \(2008年4月～2011年12月\)](#)」と同様である。ただし、2012年1月以降に提供するデータは、追加文字コードを使用せず、追加文字に相当する部分はゲタ文字 (=) に置換える。

目次

1. 使用する文字コード
2. Unicode 基本多言語面 (U+0000-FFFF) 内のコード値を持つ文字の取扱い
3. Unicode 基本多言語面 (U+0000-FFFF) 内のコード値を持たない文字の取扱い
4. 漢字の字形や意味の判断
5. 当基準の適用範囲と基準の変更に伴う訂正

1. 使用する文字コード

単行資料、逐次刊行資料ともに Unicode/UTF-8 を使用する。

2. Unicode 基本多言語面 (U+0000-FFFF) 内のコード値を持つ文字の取扱い

基本多言語面 (U+0000-FFFF) 内のコード値を持つ文字を使用する。Unicode 基本多言語面外のコード値を持つ文字については、後述 3. 「Unicode 基本多言語面 (U+0000-FFFF) 内のコード値を持たない文字」と同様に扱う。

漢字は、CJK 統合漢字 (U+4E00-9FFF) 及び CJK 互換漢字 (F900-FAFF) 内のコード値を持つ文字を使用する。

Unicode 基本多言語面内のコード値を持つ文字でも、下記①～③のいずれかに該当する場合は、当該文字を使用しない。

- ①Unicode 基本多言語面内のコード値を持つ文字が、省略しても意味上の理解が可能な装飾的な場合は、省略する。
- ②Unicode 基本多言語面内のコード値を持つ文字が、○や□で囲まれた合成文字である場合は、○や□の中の文字を ()、「 」で囲んだ形に置換える。
- ③Unicode 基本多言語面内のコード値を持つ文字を、当館のシステム上入力できなかった場合等は、後述 3. 「Unicode 基本多言語面 (U+0000-FFFF) 内のコード値を持たない文字」と同様に扱うことがある。

3. Unicode 基本多言語面 (U+0000-FFFF) 内のコード値を持たない文字の取扱い

Unicode 基本多言語面外のコード値を持つ文字、及び Unicode コード値を持たない文字は、①、②、③、④の順序で対応する。当館独自に私用面 (U+E000-F8FF) のコード値を割り当てることはしない。

- ① 意味上もしくは字形上関連する、Unicode 基本多言語面のコード値をもつ文字に置換える。
- ② 当該文字の読みを角括弧 [] で囲んだ片仮名で記録する (以下「仮名読み」という)。2 字以上の文字の音節が重なり、別個の音節を形成する連声や、熟語訓のように 1 字だけ仮名読みすると意味が不明瞭になる場合には、熟語単位で仮名読みを行う。あるいは翻字する。
- ③ 当該資料中に表示されている、「Unicode 基本多言語面のコード値をもつ文字」で表現できる別表記を、角括弧 [] で囲んで記録する。
- ④ ゲタ文字 (=) に置換える。

4. 漢字の字形や意味の判断

漢字の字形や意味の判断は、以下の漢和辞典に即して行う。

- (1) 大漢和辞典 諸橋轍次著 修訂第 2 版 大修館書店
- (2) 新字源 小川環樹 [ほか] 編 改訂版 角川書店

5. 当基準の適用範囲と基準の変更に伴う訂正

5-1. 2011 年 12 月までの記述部分のデータ

訂正しない。

5-2. 2011 年 12 月までの標目部分のデータ

字体の統一を行っていたものについては、原則として訂正しない。ただし、同一の標目が資料の情報源に違う字体で表示されている場合については、典拠データにその字体を参照形として追加する。

5-3. 2012 年 1 月以降の記述部分のデータ

当基準で記録する。

5-4. 2012 年 1 月以降の標目部分のデータ

既出の標目は、当基準に従っていないものも、原則としてそのままの形で記録する。新規の標目は当基準で記録する。

(収集・書誌調整課)