

令和時代のユニコードと ICT

永 田 善 久*

目次

はじめに	2
1 ユニコード概観	4
2 「令和合字」と Unicode 12.1.0	6
3 「ビャンビャン麺」と Unicode 13.0.0	9
3.1 文字 ID と文字マップ	12
3.2 グリフ置換機能 (GSUB)	13
4 エスツェット大文字 (U+1E9E)	15
5 フラクトゥア (ドイツ旧字体) と変種ウムラウト	18
6 かつてプリンスとして知られたアーティスト (ユニコード外の 「ラブ・シンボル」)	23
7 ウェブフォント	24
おわりに	26
参考文献	28

* 福岡大学人文学部教授

はじめに

コンピュータ上で文字を、図形データとしてではなく、テキスト形式で扱う場合、各文字には固有のバイト表現を取るコードが割り振られ、それにより各文字の一意性とデバイスを跨ぐ同一性、再現性が保証されることになるが、世界中の様々な言語環境下にあるコンピュータ間でのテキストデータの互換性を顧慮するならば、これからの「文字コード」使用は「ユニコード」一択と言っても良いように思われる。

本稿では、ここ近年におけるユニコードをめぐる動き、高度な文字組版を可能とする OpenType フォント¹の技術仕様、これらを用いたコンピュータ上での「文字組み」の可能性に関する現状を概観し、ウェブ上であれ紙媒体であれ、特に「テキストによる情報発信」という観点からの将来を見据えた「人文学」分野における現時点での ICT 利活用の可能性を探る。

本稿の趣旨から言って、本稿内で「図像」を例示し「文字組みの出力参照」を行うことはあまり意味を持たない（なぜならその時点でテキストデータではなくってしまうため）ので、代わりに OpenType フォントをユニコード（後述するが「文字符号化形式および文字符号化方式」として「UTF-8」を用いる）で取り扱う場合の「HTML5/CSS3² および TeX³ の具体的なソースコードそし

¹ アドビおよびマイクロソフトにより共同開発された OpenType フォントは、アドビの PostScript フォントを包括する形で旧来の TrueType フォントを発展拡張させたフォントファイルフォーマットである。OS を跨いで使えるクロスプラットフォーム仕様となっており、1 フォントあたり $256 \times 256 = 65,536$ 字を収録できる。特筆すべきは、「グリフ置換情報」、「グリフ位置調整情報」、「仮想ボディ情報」、といった高度な文字組版を可能とする様々な付加情報がフォントに搭載されていることである。

² FU_box に格納した各サンプルファイルに含まれる HTML/CSS のソースコードは、本稿執筆時点での最新メジャーバージョンである HyperText Markup Language のバージョン 5 および Cascading Style Sheets のバージョン 3 に準拠したシンタクス（構文規則）に従った記述にしてある。なお、HTML はウェブページの論理構造を、CSS は主に視覚デザインを記述する言語として、それぞれ役割分担させることがウェブページ作成に際しての近年の主流となっている。

³ 正確には TeX を日本語対応させた pTeX の内部ユニコード版である upTeX の拡張版の e-upTeX と、TeX をフル・ユニコード対応に拡張した XeTeX を用いる。執筆時

て出力の有様」を FU_box そして論者が運用するウェブサーバ（本学内「要塞ホスト」）を介して誰でも「直接ダウンロードまたは閲覧、そして、検証、再利用」できるようリソース先の URL を記す方式、つまり、本稿とリソース先の電子ファイル群を以て一体の論考となるような、いわばハイブリッド方式で本稿を構成する⁴。

HTML, CSS, TeX のソース（コード）ファイルは、いずれもテキストデータしか持たない「プレーンなテキストファイル」で、全て BOM (Byte Order Mark) 無しの UTF-8 でエンコーディングし、改行コードは LF (Line Feed) として作成してある。実際にこれらサンプルファイルの中身を参照したり編集したりする際には、ユニコード対応の「テキストエディタ」があると便利だろう⁵。

点における論者の TeX 環境は TeX Live 2020 に含まれる e-upTeX Version 3.14159265-p3.8.3- ul.26-191112-2.6; XeTeX Version 3.14159265-2.6-0.999992 である。TeX はフリーかつオープンソース、クロスプラットフォームのソフトウェアであるが⁶、HTML 同様マークアップ言語の一種でもあり、組版用プログラミング言語でもある。

⁴ MS Word でも「リボン」を「フォント / 詳細設定 / OpenType の機能」と辿ることで OpenType フォントが備える各種機能を一応は使えるが、論者の環境 (Microsoft Office Professional Plus 2019, Word 2019 MSO (16.0.10363.20015) 64 ビット, バージョン 1808) では、CJK 統合漢字における地域・言語間に認められる漢字表記の微細な区別を反映できない、スタイルセットを複数組み合わせで用いるユーザインタフェースが存在しない、漢字構成記述表記 (IDS: Ideographic Description Sequences) の漢字 (合字) を含むファイルを PDF 変換すると部首および漢字構成記述文字 (IDC: Ideographic Description Characters) に分解表示されてしまう、行送り量がおかしくなる、等々、様々のネガティブな現象に見舞われ、文字組みに関して OpenType フォントが本来持つ諸機能を十全に引き出すことが叶わなかったため、MS Word を使ったサンプルファイルの例示は断念した。なお、有償ソフトウェアとしては、例えば DTP ソフトウェアの Adobe InDesign が OpenType フォントに完全対応していると言われるが、論者は未確認である。

⁵ 例えば、ミニマルな機能しか持たないテキストエディタである「メモ帳」は、Windows 10 1803 (Build 17661.1001; 2018 年 5 月リリース) より前では CR (Carriage Return) LF 改行コードのみサポートしており、LF のみの改行コードを認識できなかった (LF 改行コードのテキストファイルではテキストが全く改行されずとも読みにくくなってしまっていた) が、今では改善された。LF は UNIX, Linux, Mac (OS X 以降、それより前は CR) におけるデフォルトの改行コードである (Windows でのデフォルトは CRLF)。

なお、本稿は「福岡大学情報基盤センター Web『情報の糧』⁶」に論者が寄稿した2本のレポート「No more Tofu!」(2018年5月)および「『令和』時代のユニコードと OpenType フォント⁷」(2020年5月)の続編となる論考でもある。「No more Tofu!」の中でも言及した「異体字データベース」(IVD: Ideographic Variation Database)に登録されているコレクションの一つ Adobe-Japan1-7に完全対応するフォント(23,060字のグリフより成る文字セット)として、論者のコンピュータ環境では「商用ヒラギノフォント Pr6N W3/W6⁹」を用いているが、これらの購入については「福岡大学研究推進部・領域別研究チーム『183003「応用ゲルマニステイク研究』」の研究経費に依っている。

1 ユニコード概観

コンピュータで文字を扱い、別のコンピュータと一意のデータ交換(通信)を行うには、取り扱う「文字集合」を予め定義し、その文字集合の要素である各文字を一定の規則に基づいて一意に符号化しておく必要がある。このように符号化された文字集合は「符号化文字集合」(CCS: Coded Character Set)と呼ばれるが、日本語に関しては、例えば、日本産業規格(JIS: Japanese Industrial Standards)制定のJIS X 0201(規格名称「7ビット及び8ビットの情報交換用符号化文字集合」. 現存JISにおける最古の文字コード. 1969-06-01制定)、JIS X 0208(規格名称「7ビット及び8ビットの2バイト情報交換用符号化漢字集合」. 通称「JIS 第1第2水準漢字」)、JIS X 0213(規格名称「7ビット及び8ビットの2バイト情報交換用符号化拡張漢字集合」. JIS X 0208の拡張規格。

⁶ <https://www.ipc.fukuoka-u.ac.jp/column/> (2020年9月18日閲覧)

⁷ NOTO フォント, CJK 統合漢字, 異体文字符号列, 等を扱っている。

⁸ Unicode ver. 12.1に登録された「令和合字」, CSSの各種font-feature-settingsプロパティの中でも特にdlig, trad, Character Variant等, について論じている。

⁹ macOSにバンドルされているヒラギノフォントはAdobe-Japan1-6(23,058グリフ)のサブセット(20,325グリフ)となっている。

2004年に改訂された JIS X 0213: 2004 は通称「JIS2004」と呼ばれる) 等がある。

ユニコードは、日本語表記時に通常取り扱われる文字のみならず、古今東西のあらゆる文字を共通の文字集合として利用できるようにしようという壮大な構想のもと、アドビ、アップル、グーグル、マイクロソフト等々、各種関連会社といった業界主導で制定される文字コードである。ユニコードはいわば民間の「業界規格」であり、開発の調整には非営利法人の「ユニコード協会」(The Unicode Consortium) が当たっている。もっとも、ユニコードは文字コードの公的国際規格である「ISO/IEC 10646: Universal Coded Character Set (UCS)」(International Organization for Standardization; International Electrotechnical Commission) とほぼ一致している。

さて、ユニコードの符号化文字集合に含まれる各文字には「コードポイント」(Code Point) と呼ばれる一意の数値が(例えば「a」は16進数で「61」のように¹⁰⁾振られているが、これらは(人間ではなく)コンピュータが利用できる「バイト列」(Byte String) に変換される必要があり、これは実際には「文字符号化形式」(CEF: Character Encoding Form) と「文字符号化方式」(CES: Character Encoding Scheme) という仕組みを通じて実現されている¹¹⁾。

本稿では、7ビット文字コードの ASCII (American Standard Code for

¹⁰⁾ これは通常「U+61」のように表記される (Unicode Scalar Value)。「a」のコードポイントは、10進数、8進数ではそれぞれ「97, 141」となるが、コードポイント表記に際しては、通常、0から9の数字およびAからFまでのラテン文字を使って数値を表現できることから、10進数や8進数よりも桁数が少なく済む使い勝手の良い16進数が用いられることが多い。

¹¹⁾ CEFにはUTF-8/UTF-16/UTF-32 (UTF: Unicode Transformation Format) の3種があり、CESにはUTF-8/UTF-16/UTF-16BE/UTF-16LE/UTF-32/UTF-32BE/UTF-32LE (BE: Big Endian; LE: Little Endian) の7種が定められている。なお、Big Endianという術語はコンピュータ科学者 Danny Cohen (1937-2019) が『ガリヴァー旅行記』(1726年)の「リリパット国渡航記」に登場する「卵を食する際大きな端の方を割るべきだとする連中 (Big-endians)」(カトリック宗派を揶揄していると言われる) から借用したもの。対となる Little-endian はコーエンによる造語。Big/Little Endian はバイト列を「上位/下位側から並べる」各方式を指す。スウィフト/平井『ガリヴァー旅行記』(岩波文庫, 1985) p.54, p.438 参照。

Information Interchange) の上位互換である、ASCII 文字主体の文書であれば UTF-16/UTF-32 に比べファイルサイズがぐっと小さくなる、文字境界（前後の文字バイト列との区切り）が明確でありテキスト検索時等におけるトラブルが比較的少ない、「代用対¹²」を使う必要がない、等々の理由から「1 文字を 8 ビット単位の可変長のバイト列で表現」する UTF-8 (UCS Transformation Format 8/Unicode Transformation Format-8) を、ユニコードにおける文字符号化形式及び文字符号化方式の前提とする。実際、UTF-8 はグローバルなデータ交換方式及びテキストファイル形式のデファクトスタンダードとして、現在では世界中で広く用いられている。

2 「令和合字」と Unicode 12.1.0

我が国における 2019 年 5 月 1 日の改元と歩調を合わせるように、ユニコード協会は 2019 年 5 月 7 日に「令和合字」(SQUARE ERA NAME REIWA) を含む Unicode version 12.1 を公開した（総数 137,929 文字）。「令和合字」はこのために予め宛がわれていたコードポイント U+32FF (16 進数) を持つが、ユーザがこの文字を実際に使えるかどうかは、使用フォントが当該グリフ（字形）を持っているかどうかによって依存する。例えばヒラギノフォントは「横組み用・縦組み用」の「2 グリフ」を備えている。

ユニコードにおける正式なバージョンは version 12.1.0 のように「メジャーバージョン、マイナーバージョン、アップデートバージョン」の順に数値で記されるが、マイナー／アップデートバージョンの箇所が 0 の場合はこれらの数値が省略されることもある¹³。

コードポイントの他に、各ユニコード文字は一種の「学名」に相当する「(英

¹² Surrogate Pair. UTF-16 で用いられる符号化の拡張方法。2 バイトで定義される基本多言語面 (BMP: Basic Multilingual Plane: 65,536 文字分の符号領域) に含まれない拡張領域の文字を、2 バイトのバイト列 2 つを組み合わせることで表す方式。

¹³ 因みに version 12.0 から 12.1 への改訂は、実に「令和合字 1 字」のみの追加であった。

語の) 公式文字名」(Formal Character Name) を持ち、これは全て「大文字」で記される(例: LATIN SMALL LETTER A)。例えば GNU Emacs のような「フル・ユニコード対応」のテキストエディタであれば、16 進数のコードポイントあるいは公式文字名を与えることで全てのユニコード文字を入力できる¹⁴。

以下、本稿では

HTML5/CSS3 のサンプルファイル

TeX のソースファイルおよびコンパイル¹⁵ 済みの PDF ファイル

に関する URL を記すので、関心のある向きはリソース先の各種サンプルファイルを直接参照、検証、そして、自由に再利用されたい。

各種サンプルファイルについては、自己完結した(出来得る限り)ミニマルなサンプルを提供することに主眼を置いた。例えば、HTML と CSS はそれぞれ別ファイルとせず、CSS のソースコードを HTML の head 要素 (Element) 内の style 要素中に「埋め込む」方式で「CSS 記述部を含む 1 つの HTML ファイル」として作成した(従って、この記述部のシンタクスは HTML ではなく CSS に則っている)。

¹⁴ GNU Emacs 上で C-x 8 RET (ユニコード文字入力のためのキーバインディング) するとミニバッファで Insert character (Unicode name or hex): と尋ねてくるので、また、公式文字名は「入力補完」対象となっているため、簡単かつインタラクティブにユニコード文字の入力が可能である。もちろん、当該文字が実際に出力されるかどうかはユーザの用いるフォントがそのユニコード文字をサポートしているかどうかによる。Emacs では、他にも、当該文字のユニコードにおけるコードポイント、公式文字名、ユーザが使っている文字集合におけるコードポイント、当該文字のキーボードからの入力方法(キーバインディング)、ファイル中の当該文字のエンコーディング、等々といった様々な文字情報を C-u C-x = RET で即座に参照できる。

¹⁵ ここではプレーンテキストである「TeX のソースコード」をバイナリである PDF という「オブジェクトコード」に変換する工程を、プログラミング言語における術語になぞらえて、このように表現している。

サンプルファイル群は、全て、FU_box 内「外部共有」セクションに設置した 1 フォルダ、具体的には「リンクを知っている全員」に対して「表示およびダウンロード可能」属性を与えた 1 フォルダ

<https://fukuoka-u.box.com/s/rlohnd2257gh3d0lvzg7qss16vsnf8o>

の中に、ファイル名を一意にして、格納してある。以下、FU_box に格納したサンプルファイル名をリストに挙げる際は、重複を避けるため上の URL 表記は繰り返さない。

HTML ファイルについては、論者運用のウェブサーバを介してウェブ上でも閲覧可能としてある（その都度 URL を記してある）。ただし、後述する「ウェブフォント」の提供がない限り、閲覧者のシステムに当該グリフ（字形）を含む OpenType フォントが既にインストールされていなければ論者が意図した「十全な見え方」とはならないため、比較参照用に論者の環境におけるウェブブラウザ閲覧時画像も HTML 内に添えておいた。全 HTML サンプルファイルは論者により、Microsoft Edge (ver. 85.0.564.44), Firefox (ver. 80.0.1), Google Chrome (ver. 85.0.4183.83), Safari (ver. 13.1.2 15609.3.5.1.3) での閲覧検証を経ている。なお、各種 HTML サンプルファイルの Safari での出力は色々と瑕疵が認められた¹⁶ ため、Mac 上でのサンプル HTML の閲覧に際しては Firefox 等のブラウザを利用していただけると有難い。

TeX ファイルに関しては、ウェブサーバ上で TeX ソースファイルを PDF

¹⁶ Safari では後述する GSUB の「機能識別子」pwid, palt, pkna, fwid, hwid, halt, twid, qwid, jp78, jp83, jp90, jp04, trad, hkna, nlck, nalt, ital, vkrn, vpal, vert, vpal, vhal, vkna がサポートされていない。Adobe: *Syntax for OpenType features in CSS*. <https://helpx.adobe.com/fonts/user-guide.html/fonts/using/open-type-syntax.ug.html> (2020 年 9 月 18 日閲覧)

へコンパイルできるウェブアプリケーション¹⁷

[upTeX TYPESET: upTeX でコンパイル]

https://lgs.hum.fukuoka-u.ac.jp/~ynagata/uptex_typeset/

[XeTeX TYPESET: XeTeX でコンパイル]

https://lgs.hum.fukuoka-u.ac.jp/~ynagata/xetex_typeset/

を作成・設置しておいたので、TeX システムを持たないユーザであってもウェブ上で自らサンプルファイルをコンパイルし、追試することが可能である。その際 upTeX/XeTeX のどちらを用いるのか、については、サンプルファイル横の [] の中に、使用エンジンの別を記しておいた。

[令和合字 (横組み・縦組み) : FU_box (ダウンロード用)]

01_square_era_name_reiwa.html

01_square_era_name_reiwa.tex [upTeX]

01_square_era_name_reiwa.pdf

[令和合字 (横組み・縦組み) : Web (ウェブ閲覧用)]

<https://lgs.hum.fukuoka-u.ac.jp/~ynagata/2020/jinbun-ronso/>

01_square_era_name_reiwa.html

3 「ビャンビャン麺」と Unicode 13.0.0

2020年3月10日に公開されたユニコード・バージョン13は、前バージョンの12.1から5,930字増となった「総文字数143,859字」を規定している。

これら増分の中で、テキスト(文字)処理に係わる ICT スキルとの関連に

¹⁷ 本ウェブアプリケーションは Linux, Apache, PHP, TeX を組み合わせて論者がプログラムした。

において特に取り上げたい文字の一つが「ビャン」である。中国の陝西省で食される幅広の麺の名称（「ビャンビャン麺」）中に含まれるこの文字は、簡体字で43画（コードポイントはU+30EDD）、繁体字で58画（コードポイントはU+30EDE）の画数を持つ複雑な文字で、いずれもユニコード・バージョン13から他の4,937の漢字とともに第三漢字面（TIP: Tertiary Ideographic Plane）にCJK（Chinese/Japanese/Korean）統合漢字拡張G（CJK Unified Ideographs Extension G）として収録された¹⁸。

論者が調べた限りでは、本稿執筆の時点でこの「ビャン」文字をグリフとして含むOpenTypeフォントは、2017年5月8日にリリースされたSource Han Serif (Version 1.001)と、2019年4月9日に改訂リリースされたSource Han Sans (Version 2.001)の二つのフォントファミリーのみである。これらは、アドビとグーグルがオープンソースとして共同開発した初めての巨大な「汎中日韓」(Pan-CJK)フォントファミリーで、統一された書体デザインと7つのウェイト（文字の太さ）を持つ。日本語ではそれぞれ「源ノ明朝；源ノ角ゴシック」と呼ばれるが、例えば中国語では「思源宋体、思源宋體；思源黒体、思源黒體、思源黒體香港」といった具合に「ほぼ同じフォント」であるにも拘らずそれぞれ名

¹⁸ The Unicode Consortium: *The Unicode Standard. Version 13.0 – Core Specification*. <http://www.unicode.org/versions/Unicode13.0.0/UnicodeStandard-13.0.pdf> (2020年9月18日閲覧) ユニコードの全領域は、0-16面 (Plane), 0-255区 (Row), 0-233点 (Cell) の「 $17 \times 256 \times 256 = 1,114,112$ 」個から成る符号位置によって表される。各面には $256 \times 256 = 65,536$ 個の文字を収録し得るから、実際、「基本多言語面」(BMP: Basic Multilingual Plane) と呼ばれる第0面には、世界で通常用いられる基本的な文字のほとんどが入っている。これを超える第1面「追加多言語面」(SMP: Supplementary Multilingual Plane) には線文字Bやゴート文字といった古代のスク립トや音楽記号等が含まれ、第2面「追加漢字面」(SIP: Supplementary Ideographic Plane) にはBMPから漏れた漢字が収録されており（例えば日本語で用いられる漢字の異体字等もこの面に入っている）、第3面「第三漢字面」にはSIPに入りきらなかった漢字や甲骨文字・金文・小篆といった古代漢字が収録される予定となっている。残りの第4から13面については現時点で用途未定、第14面は「追加特殊用途面」(SSP: Supplementary Special-purpose Plane)、第15及び16面は「私有面A及びB」(PUA-A/B: Private Use Area-A/B Plane) である。

称が異なる。実際、同じコードポイントを持つ漢字でも中国語（簡体字）、中国語（繁体字）、日本語、韓国語でグリフ（字形）の細部が異なるデザインとなっている。なお、同じフォントであっても、グーグルでは巨大な Noto (No more Tofu) フォントファミリの一部として Noto Serif/Sans CJK という名称が付けられている。

Source Han Serif/Sans はユーザが使用している OS 等に応じて、以下の 4 つの配布形態

Language-specific OpenType/CFE (OTF)

OpenType/CFE Collection (OTC)

Super OpenType/CFE Collection (Super OTC)

Region-specific Subset OpenType/CFE (Subset OTF)

で提供される。論者の環境（Windows 10 1909-OS Build 18363.1049; macOS Catalina 10.15.6）では、Windows および Mac のいずれにおいても Super OTC（全言語文字を 1 ファイルに収容）をインストールしている¹⁹。Source Han Serif/Sans は SIL (Summer Institute of Linguistics) Open Font License 1.1 の規約に則り、フリーフォントとして、ソースコード（ソフトウェア）開発のプラットフォーム GitHub から自由にダウンロードできる。GitHub に不慣れであれば、「公式ダウンロードガイド」を参照すると良い。日本語による説明もある。

<https://github.com/adobe-fonts/source-han-serif> (2020 年 9 月

¹⁹ Windows 上の MS Word では現時点で Super OTC がサポートされていないので、Word ユーザは OTC（7 ファイルに分割して全言語文字収容）あるいは OTF（全言語文字を言語別の 4 サブディレクトリ、28 ファイルに分割して収容）を選択する方が良い。Mac 上の MS Word について論者は未確認。

18 日閲覧)

<https://github.com/adobe-fonts/source-han-sans> (2020 年 9 月 18 日閲覧)

<https://github.com/adobe-fonts/source-han-serif/raw/release/download-guide-source-han.pdf> (2020 年 9 月 18 日閲覧)

3.1 文字 ID と文字マップ

フォントが収録する個々の文字には、当該フォントに固有の管理識別用「文字 ID」(GID: Glyph ID/CID: Character ID) が振られている。文字 ID は文字コードとはひとまず無関係であるから、あるフォントを実際に使おうとする場合、当該フォントの各文字 ID と文字コードを一意に結び付ける対応表が必要となる。この対応表に相当する仕組みが「文字マップ」(CMap/cmap) である。OS やアプリケーションソフトウェアは文字マップを介して必要とする文字 ID を突き止め、フォントの各文字にアクセスする²⁰。

さて、ビャンビャン麵の「ビャン」のグリフを持つ OpenType フォント Source Han Serif (Version 1.001)/Sans (Version 2.001) は、それぞれ 2017 年 5 月 8 日及び 2019 年 4 月 9 日にリリースされたため、2020 年 3 月 10 日公開のユニコード 13 で初めて収録された文字に関する文字マップを、当然、備えていない。従って、文字コード・文字マップ・文字 ID を連携させた当該文字へのアクセスは不可能である。では、Source Han Serif/Sans に含まれる「ビャン」の字をコンピュータで取り扱うには、一体、どのようにしたら良いのか。

²⁰ TrueType および OpenType フォントでは GID と cmap を、PostScript 処理系のフォントでは CID と CMap を用いる。例えば TeX の OTF パッケージのように、アプリケーションソフトウェアによっては CID 番号を直接操作し、フォントの全文字にアクセスできるものもある。

3.2 グリフ置換機能 (GSUB)

こうした場合に用いられる仕組みが、OpenType フォントが備える様々な機構のうちの一つ「グリフ置換機能」(GSUB: Glyph SUBstitution feature) である。

GSUB は「英数字 4 字」から成る「機能識別子」(Feature Tag)、例えば、frac (斜線分数置換)、dlig (活版印刷用合字置換)、liga (欧文標準合字置換)、ruby (ルビ用かな置換)、vert (縦組用置換)、jp78 (JIS78 字形置換)、jp83 (JIS83 字形置換)、jp04 (JIS2004 字形置換)、trad (旧字体置換) 等々、を用いることにより高度な文字組版を可能とする機能である。「ビャン」字の場合は、こうしたグリフ置換機能のうち「ccmp (glyph Compositon/deCoMPosition: 合字/分解置換)」を用いれば良い。一般的に GSUB の実装は各 OpenType フォントにより異なるが、Source Han Serif/Sans は ccmp をサポートしている。

具体的には、ソースコード上で「漢字構成記述文字」(IDC: Ideographic Description Character) を用いた「漢字構成記述表記」(IDS: Ideographic Description Sequence) による入力を行うが、詳しくはサンプルファイル内の当該記述箇所を参照されたい²¹。

HTML ファイルとしては、html 要素の lang 属性を使って「日本語、韓国語、簡体字中国語、繁体字中国語香港、繁体字中国語台湾」を「言語指定」することにより、5つの異なるサンプルを作成した。「神は細部に宿る²²」。「ビャン」の各グリフにおける細かな差異をじっくりと比較されたい。TeX(XeTeX)ファイルでは「ビャン」の全グリフを「A4 版 1 枚のチェックシート」にまとめたので、まずこちらを概観すると各グリフの違いを認識しやすくなるかも知れない。

²¹ 「漢字構成記述表記」については *Source Han Sans Version 2.001* (Built on April 4, 2019; Released on April 9, 2019) も参照。 <https://github.com/adobe-fonts/source-han-sans/raw/release/SourceHanSansReadMe.pdf> (2020 年 9 月 18 日閲覧)

²² Der liebe Gott steckt im Detail. 美術史家アビ・ヴァールブルク (1866–1929) によるとされる表現。

ccmp の別サンプルとして、同様にユニコード・バージョン 13 から登録された「たいと」(U+3106C; 総画数 84 画の和製漢字)と「三水に恩」(U+30729; 韓国の人名用漢字らしいが論者は無案内)も挙げておいた。これらの文字も Source Han Sans には含まれている (が Source Han Serif には収録されていない)。

なお、論者の Mac (ver. 10.15.6) のブラウザ Safari (ver. 13.1.2) では「ビャン」字の合字置換が機能しなかった。

[ビャンビャン麵：FU_box (ダウンロード用)]

02_biang_biang_ja.html

02_biang_biang_ko.html

02_biang_biang_zh-cn.html

02_biang_biang_zh-hk.html

02_biang_biang_zh-tw.html

02_biang_biang.tex [XeTeX]

02_biang_biang.pdf

[ビャンビャン麵：Web (ウェブ閲覧用)]

<https://lgs.hum.fukuoka-u.ac.jp/~ynagata/2020/jinbun-ronso/>

02_biang_biang_ja.html (日本語)

<https://lgs.hum.fukuoka-u.ac.jp/~ynagata/2020/jinbun-ronso/>

02_biang_biang_ko.html (韓国語)

<https://lgs.hum.fukuoka-u.ac.jp/~ynagata/2020/jinbun-ronso/>

02_biang_biang_zh-cn.html (簡体字中国語)

<https://lgs.hum.fukuoka-u.ac.jp/~ynagata/2020/jinbun-ronso/>

02_biang_biang_zh-hk.html (繁体字中国語香港)

<https://lgs.hum.fukuoka-u.ac.jp/~ynagata/2020/jinbun-ronso/>

02_biang_biang_zh-tw.html (繁体字中国語台湾)

ユニコード・バージョン 13 には、他にも例えば「鏡」を 4 つ「縦横 2 行 2 列」の正方形に並べた漢字 (U+30F54) も登録されたが、この文字は Source Han Serif/Sans フォントには (残念ながら) 含まれていない²³。

4 エスツェット大文字 (U+1E9E)

ドイツ語アルファベットで用いられる「エスツェット (sz) B」は、本来「長い s」(次節「フラクトゥア (ドイツ旧字体) と変種ウムラウト」も参照) と「尻尾の付いた z」との合字である。

長い s の大文字が存在しない以上、sz の「大文字」合字というのは不合理な気がしないでもないが、民間では例えば標識や店の看板等で、いわば「エスツェット」を一文字と捉え「エスツェット大文字」を柔軟に用いてきたという現実があった。

2008 年 4 月 4 日、紆余曲折を経て、エスツェット (B; それまでは小文字のみが存在) の対となる「エスツェット大文字」がコードポイント「U+1E9E」としてユニコード・バージョン 5.1²⁴ に取り入れられることとなった (ISO/IEC 10646 には 2008 年 6 月 24 日)。その後、各方面での慎重な検討および 9 年の

²³ この漢字は、宮沢賢治が「岩手軽便鉄道の一月」(1926 年) という詩の中で実際に用いている。草野心平編『宮沢賢治詩集』(新潮文庫, 1979 年) p.118 参照。賢治は「みなまっ白に凍ってゐる」「河岸の樹」に向けて、それぞれ学名調の名称でもって一つ一つ親しく呼びかけていくのだが、その際、恐らくは凍てついた各樹果が太陽光をキラキラと反射させる様をそれぞれに「鏡をつるし」と単数描写した。しかし、「はんのきアルヌスランダー」に対してだけは「U+30F54」文字を使ったのである。これは、恐らく、ハンノキ (*Alnus japonica*) の松かさ状の堅果が複数 (4 から 5 個ほど) 群れてなる様を賢治が正確に描写したかったためだと思われる。因みに「森の妖精」あるいは「森の宝石」とも呼ばれる「ミドリシジミ」(*Neozephyrus japonicus*) の幼虫はハンノキを食草の一つとする。

²⁴ <http://www.unicode.org/versions/Unicode5.1.0/> (2020 年 9 月 18 日閲覧)

歳月を経てようやく、2017年6月29日、「ドイツ語正書法評議会²⁵」(RdR: Rat für deutsche Rechtschreibung)は「エスツェットの太文字表記法として従来のSSと並んで新たにエスツェット太文字も認める」という内容が含まれる声明を発表した²⁶。この時点から、「エスツェット太文字」は「ドイツ語正書法に係る公認規定」(Die amtliche Regelung der deutschen Rechtschreibung²⁷)の一部となった。従って学校教育で教授されねばならないといった「法的拘束力」を持つ。

ただし、「エスツェット太文字」(U+1E9E)は一般的な欧文フォントにあたりまえに収録されている文字では、少なくとも現時点では、ない。ユニコードでは「基本ラテン文字」(Basic Latin)のブロックに含まれるエスツェット(U+00DF)とは異なり、同じ第0面「基本多言語面」の中ではあるものの、「ラテン文字拡張追加」(Latin Extended Additional²⁸)のブロックの中に分類されているのである。

以下のサンプルファイルでは「エスツェット太文字」を含むOpenTypeフォントとしてフリーフォント(Open Font License, ver. 1.1に従う)のLibertinus Sansを用いている。LibertinusはLinux Libertineフォントから枝分かれ(fork)し開発が進められている(特に数式用フォントの追加)フォントプロジェクトによる成果物で、Linux Libertine同様「ラテン文字、ギリシア文字、キリル文字、ヘブライ文字」間のデザイン統一(Pan-Language Harmony)が特徴で

²⁵ 2004年にドイツ、オーストリア、スイス、ボルツァーノ自治県(「南ティロル」とも)、リヒテンシュタイン、ベルギードイツ語共同体により設立された「ドイツ語正書法」の策定に携わる国際機関。本部はドイツ国マンハイムに置かれている(IDS: Leibniz-Institut für Deutsche Sprache内)。

²⁶ https://www.rechtschreibrat.com/DOX/rfdr_PM_2017-06-29_Aktualisierung_Regelwerk.pdf (2020年9月18日閲覧) これ以外にも、例えばGoldene Hochzeit(金婚式)とかNeues Jahr(新年)といったような慣用表現の表記時に「形容詞の頭字太文字書き」を認める、といった事柄も含まれる。

²⁷ https://www.rechtschreibrat.com/DOX/rfdr_Regeln_2016_redigiert_2018.pdf (2020年9月18日閲覧)

²⁸ <https://www.unicode.org/charts/PDF/U1E00.pdf> (2020年9月18日閲覧)

ある。Libertinus Serif/Sans は GitHub から自由にダウンロードして利用できる。

<https://github.com/alерque/libertinus> (2020 年 9 月 18 日閲覧)

以下の HTML サンプルファイルでは、「エスツェット大文字」を直接表記および 16 進数数値文字参照表記の 2 通りで入力してある。また、css の `text-transform` プロパティの値を `uppercase` とすることで、小文字で入力した文字列が大文字として出力されるスタイル指定を施した箇所と、`font-variant` プロパティ値を `small-caps` とした箇所（当該箇所がスモールキャピタルで出力される）がある。前者では、HTML のオリジナル仕様によりエスツェット（小文字）は「SS」として出力される。

オリジナル仕様の TeX (XeTeX) でも、エスツェット（小文字）は大文字変換される際「SS」として出力されるようにプログラムされているが、サンプルファイル内ではこの仕様を少々書き換えることで、「エスツェット大文字を含むフォントとそうでないフォント」それぞれの使用時における出力が「エスツェット大文字または SS」となるような仕掛けを施しておいた（ソースファイルの「プリアンブル」（前口上）部 7, 8 行目を参照）。また、macOS で XeTeX を用いる場合、「macOS の XeTeX では TeX のシステムフォントをフォント名では呼び出せない」（Linux, Windows であれば可能）という制約があるため、フォント名の代わりに「フォントファイル名」をソースに記す必要がある。サンプルファイルでは Mac 用の記述を「コメントアウトした形で 11, 12 行目」に記しておいた。Mac を使う場合は前行（9, 10 行目）の代わりにこちらの行の記述を生かされたい。

[エスツェット大文字：FU_box (ダウンロード用)]

03_latin_capital_letter_sharp_s.html

03_latin_capital_letter_sharp_s.tex [XeTeX]

03_latin_capital_letter_sharp_s.pdf

[エスツェット大文字：Web（ウェブ閲覧用）]

<https://lgs.hum.fukuoka-u.ac.jp/~ynagata/2020/jinbun-ronso/>

03_latin_capital_letter_sharp_s.html

5 フラクトゥア（ドイツ旧字体）と変種ウムラウト

ドイツ語繋がりとして、本節ではフラクトゥア（ドイツ旧字体）と変種ウムラウトをユニコードおよびOpenType フォント関連トピックとして取り上げる。

活版印刷が普及する前の中世手写本の時代においては、ウムラウトには様々な表記法があった（詳細は割愛）。西洋初の活版印刷による『ゲーテンベルク42行聖書』は、「カロリング朝小文字体」（karolingische Minuskel）の後に登場した手書きの「ゴシック小文字体」（gotische Minuskel）から発達した「テキスト書体」（Textura）の一種を用いているが、これは「フラクトゥア」（Fraktur; 原義は「折れ曲がった書体」）の一種に分類される。フラクトゥアはこうしたテキスト書体から形作られてくるが、その発展は（ブルゴーニュ公、オーストリア大公、ドイツ王、神聖ローマ皇帝の）マクシミリアンI世と密接に結びついている（詳細は割愛）。

活版印刷におけるフラクトゥアでウムラウトを表記する際は「字母の後に e を送る、小さな e を字母の上に添える²⁹」といった方法が取られたが、特に字母

²⁹ 現在では「2点」式のウムラウト表記となっているが、なぜ小添字 e 式から2点式へと代わっていったのか、はフラクトゥアの「筆記体」（Kurrent; 原義は「走る、流れる」）の e の字を（羽根ペンで、と言いたいところだが）スタブ・ニブの万年筆（斜線は細く、縦線は太く描かれる）で書いてみれば、その理由がすぐに分かる。フラクトゥアの筆記体による e はラテン文字 e とは似ても似つかない、慣れないと n と誤読してしまうような、

が大文字の場合、鉛活字の上部に e を組み込む十分なスペースが確保しにくい
ため、「e 後置方式」で組まれることが一般的であった³⁰。

さて、ユニコードの下、フラクトゥアをコンピュータで「精緻かつ正確に」
扱おうとする場合、最新のフォントテクノロジーを備えた OpenType を用い
る必要がある。というのは、ユニコードでは「通常の欧文テキストで用いられ
るフラクトゥア」は独立した文字ではなく、あくまで「ラテン文字の文字変種」
として捉えられているのであるが、実際にはフラクトゥアにはラテン文字には
見られない数多くの合字（さらに、隔字で組まれるときは合字が解除される場
合とそうでない場合がある、等々）その他の複雑な組版慣習が伴っているので、
これらを正確に再現するには、OpenType フォントが持つ高度な文字組版を
可能とする様々な付加情報がどうしても必要となるためである。

ここで、若干、事前の整理が必要となる事柄もある。それは、通常の欧文テ
キストで用いられるフラクトゥア「以外」の用途で用いられるフラクトゥア文
字も存在する、という点だ。例えば数学における「リー代数」(Lie Algebra)
はフラクトゥアの小文字で表される。「 σ -集合代数」(σ -Algebra)、「イデアル」
(Ideal)、ラマヌジャン和 (Ramanujan Sum) 等の表記にもフラクトゥアが使
われることがある。こうした事情を反映し、ユニコードの第 1 面「追加多言語
面」には「数学用英数字記号³¹」(Mathematical Alphanumeric Symbols)のブロッ
クが設けられており、ここに「数学用フラクトゥア大文字 A」(U+1D504;
MATHEMATICAL FRAKTUR CAPITAL A) 以下が登録されている（小文字
もあれば太字もある、例えば U+1D56C: MATHEMATICAL BOLD FRAKTUR
CAPITAL A; U+1D586: MATHEMATICAL BOLD FRAKTUR SMALL A)。

字形である。

³⁰ なお、「ドイツ語正書法に係る公認規定」の中にフラクトゥアに関する取り決めは皆
無である。従って、大文字字母ウムラウトにおける「e 後置方式」は現在では拘束力
を持つルールというよりも、一種の組版慣習と言える。

³¹ <https://unicode.org/charts/PDF/U1D400.pdf> (2020 年 9 月 18 日閲覧)

ところが、具に「数学用英数字記号」ブロックをチェックしていくと、フラクトゥア大文字のうち「C, H, I, R, Z」が欠けている（これらのコードポイントは将来用に「予約」(reserved) となっている）ことが分かる。実は、これら5文字は第0面「基本多言語面」の「文字様記号³²」(Letterlike Symbols) に存在している。なぜこのようになっているのか、たとえば、フラクトゥア大文字「C, H, I, R, Z」の5文字はユニコード1.0 (1991年10月制定) で、それ以外のフラクトゥア大文字はユニコード3.1 (2001年3月制定) で、それぞれ別時期に登録されたという経緯があるためだ。当該5文字には昇順に「U+212D: 術語としてのフラクトゥアCが何を表すのかは論者には不明³³, U+210C: ヒルベルト空間を表す, U+2111: 複素数虚部を表す, U+211C: 複素数実部を表す, U+2128: 薬衡ドラム記号 (1/8 オンス)」といったコードポイントが付与されている。

「数学用英数字記号」であれ「文字様記号」であれ、これらのブロックに登録されたフラクトゥアにはウムラウトも、「長いs」(U+17F) も、各種合字の類も、一切付随していない。従って欧文テキスト組版用には使えない。ギリシア文字「シグマ」の使い分けと同じように、フラクトゥアにおける長いsは「語頭および語中」で用いられ、丸いsは「語尾、そして語中にあってもそれが合成語である場合は語区切りの語尾」で用いられた。

因みに長いsは、フラクトゥアのみならず、かつてはラテン文字でも使われていた。例えば『ガリヴァー旅行記』初版(1726年)の中での挿絵地図の中では「Blefuscu Lilliput. Discovered, A.D.1699」とある語中のsがいずれも長いsとなっていることが確認できる³⁴。

³² <https://unicode.org/charts/PDF/U2100.pdf> (2020年9月18日閲覧)

³³ HTMLの「文字実体参照」(Character Entity Reference)ではℭと記すので、ひょっとすると「ケイリー・ハミルトンの定理」で知られるアーサー・ケイリー(1821-1895)に関連するのかも知れない。

³⁴ Map of Lilliput and Blefuscu: <https://www.bl.uk/collection-items/map-of->

以下のサンプルファイルでは、フラクトゥアの OpenType フォントとして「UnifrakturMaguntia³⁵」を使用している。UnifrakturMaguntia は Open Font License 1.1 に則り、フリーフォントとして、以下のウェブサイトから自由にダウンロードして利用できる。

<http://unifraktur.sourceforge.net/maguntia.html> (2020 年 9 月 18 日閲覧)

HTML ソースファイルでは「ウムラウトは通常の 2 点式で、長い／丸い s は入力した通りに」出力される例を、先ず、示しておいた。次に、CSS の「スタイル・セット」(SS: Stylistic Set) の ss11 (長い s と丸い s を自動的に組み分ける), ss13 (大文字の I は J で出力する), ss15 (小添字 e 式ウムラウト), ss14 (大文字に関しては小添字 e 式ウムラウトとせず大文字の後に e を送って表記する) を活かすことで、そのような出力が「長い／丸い s の入力如何に係わらず同様に実現される」例を示した。このことは「入力段階で全て丸い s で入力しておいても、出力時には自動的に長い／丸い s に組み分けられる」ことを意味するから、例えば時代考証を反映させたフラクトゥアのテキストを組む際等に、この恩恵は大きい。最後に、「隔字体」(Sperrsatz; letter spacing) の例も挙げた。OpenType フォントの UnifrakturMaguntia では、「技術」的にはもちろん「イタリック体」(一種の斜字体) を用いることが可能である。しかし、イタリックとは本来「イタリア風」の謂であるから、「ドイツ旧字体 (フラクトゥア) のイタリア風」書体はそもそも形容矛盾であり、実際フラクトゥ

lilliput-and-blefuscu (2020 年 9 月 18 日確認)

³⁵ 1901 年 Carl Albert Fahrenwaldt によってデザインされた Mainzer Fraktur を Peter Wiegel がデジタル化した Berthold Mainzer Fraktur を基に Gerrit Ansmann によってフォントが OpenType 化されたものが UnifrakturMaguntia である。なお Maguntia は Mainz のラテン語名。

アが斜字体で用いられることはなく、強調箇所は別のフラクトゥア書体 (Schwabacher 等) あるいは隔字体で組まれてきた。しかも、エスツェット (1 字と見なされる) は言うまでもなく、「ch, ck, (st,) tz の 3 つ (4 つ) の合字は隔字体となった場合でも合字解除されない」という組版慣習³⁶があった。UnifrakturMaguntia はこうした細かな慣習も反映させた出力を自動的に行うので、入出力を注意深く比較されたい。比較参照用として「ラテン文字における小添字 e 式ウムラウト」(U+364: COMBINING LATIN SMALL LETTER E) と「長い s」(U+17F) を用いた例も付けておいた³⁷。CSS で uppercase や small-caps 指定した際の挙動も確認できる。なお、ラテン文字における小添字 e 式ウムラウトはイタリックやスモールキャピタルにも追従する。

TeX サンプルファイルも基本的に HTML ファイルと同様の内容がチェックできる構成としてある。両者の出力について言及しておく、HTML の隔字体箇所では「e 後置式大文字ウムラウト」はあたかも一文字であるかのように隔字されなかったが、TeX では隔字される。

[フラクトゥアと変種ウムラウト, 長い s : FU_box (ダウンロード用)]

04_fraktur_umlaut.html

04_fraktur_umlaut.tex [XeTeX]

04_fraktur_umlaut.pdf

[フラクトゥアと変種ウムラウト, 長い s : Web (ウェブ閲覧用)]

³⁶ ドイツ語の正書法辞典として著名な Duden では、ここに st を含めていない。*Duden Band 1. Die deutsche Rechtschreibung*, 22., völlig neu bearbeitete und erweiterte Auflage. Hrsg. von der Dudenredaktion. Auf der Grundlage der neuen amtlichen Rechtschreibregeln. Mannheim/Leipzig/Wien/Zürich 2001, S. 96 f. 参照。

³⁷ 古い文学作品におけるオリジナルのフラクトゥア組版をラテン文字組みに翻刻することがあるが、その際「小添字 e 式ウムラウトまでそのまま翻字」される場合もある。*Das Lalebuch*. Nach dem Druck von 1597 mit den Abweichungen des Schiltbürgerbuchs von 1598 und zwölf Holzschnitten von 1680. Hrsg. von Stefan Ertz, Philipp Reclam Jun. Stuttgart, 1982. 参照。

https://lgs.hum.fukuoka-u.ac.jp/~ynagata/2020/jinbun-ronso/04_fraktur_umlaut.html

6 かつてプリンスとして知られたアーティスト (ユニコード外の「ラブ・シンボル」)

アルバム・シングル総売り上げが1億3000万枚以上と言われるアメリカの音楽家プリンス (Prince Rogers Nelson; 1958–2016) は、1990年代、契約会社であるワーナー・ブラザーズとの確執から音楽活動 (ビジネス) において「プリンス」の名を用いることを止め、代わりに「どのように発音して良いのか分からない (unpronounceable) シンボル」 (通常「ラブ・シンボル」と呼ばれる) を自らの芸名とした。日常使用において「発音不能」な芸名は、当然、不便であるから、Prince Rogers Nelson は実際には「かつてプリンスとして知られたアーティスト」 (the artist formerly known as Prince あるいは単に the Artist) として自ら名乗ったり人から呼ばれたりした。

この「ラブ・シンボル」は、現時点では、ユニコードに登録されていない。が、1994年に発売された「Prince Interactive」というコンピュータ・ゲームの中に当該シンボルを含むフォント (ただし、現在の PC では使えないフォーマット) を見つけ出した人物 (Mike Wehner³⁸) がおり、彼によりそのフォントは TTF フォントへとフォーマット変換され、以下の URL

<https://www.dropbox.com/sh/jvj2ss4jfyvmvir1/AADXW3RWGwgykjlmgSInvCRBa?dl=0> (2020年9月18日閲覧)

より PrincePlain.ttf というファイル名でダウンロード可能 (フリー) となっ

³⁸ <https://www.dailydot.com/debug/prince-font-install-love-symbol/> (2020年9月18日閲覧)

いる（使用は自己責任で）。PrincePlain.ttf は「大文字 S」に「ラブ・シンボル」をマッピングしている。従ってこのフォントの使用時に「大文字 S」を用いると「Sではなくラブ・シンボルが出力される」³⁹。

[ラブ・シンボル：FU_box（ダウンロード用）]

05_lovesymbol.html

05_lovesymbol.tex [XeTeX]

05_lovesymbol.pdf

[ラブ・シンボル：Web（ウェブ閲覧用）]

<https://lgs.hum.fukuoka-u.ac.jp/~ynagata/2020/jinbun-ronso/>

05_lovesymbol.html

7 ウェブフォント

本稿で扱った Source Han, Libertinus, UnifrakturMaguntia, PrincePlain といったフォントは、ある意味では特殊用途のフォントであり、一般的なユーザの PC にインストールされる（されている）ことは稀であるかも知れない。

そもそもウェブブラウザはユーザのコンピュータ端末にインストールされているフォントを呼び出して文字を表示するので、当該フォントが存在しなければいわゆる「トーフ（豆腐）」（本来呼び出されるべき文字の代わりに表示される四角いボックス）が出現したり、あるいは情報発信者側の意図した文字とは異なる文字で代替表示されてしまったりする。

こうした問題を解消するために開発された技術が「ウェブフォント」（Web Typography; Web Font）である。ウェブフォントとはウェブサーバ側がフォントデータを持ち、これをユーザ（クライアント）側へ送信することで、「特

³⁹ 論者の Mac におけるブラウザ環境では、Safari でも Firefox でも「ラブ・シンボル」は出力されなかった。原因不明。

殊フォントを持たないユーザの端末環境でも情報発信者が意図した文字表示がなされる」仕組み、また、この用途のためのフォント形式 (eot: Embedded OpenType; woff: Web Open Font Format; woff2 等) のことを言う。ウェブフォントは CSS3 の「フォント・モジュール」(CSS Fonts Module) によって技術仕様の標準が定められている⁴⁰。

ウェブフォントは便利な反面、実際の使用に際しては注意も必要である。というのは、情報発信者が多数の漢字を含むようなフォントをウェブフォントとして用いれば、必然的にファイルサイズの大きなフォントデータのダウンロードをウェブページ閲覧者 (ユーザ側) に強いることになってしまう、という短所も持ち合わせているからである。ファイルサイズが大きければユーザ端末での文字表示 (レンダリング) までの時間も、当然、長くなる。フォントのライセンスにも気をつけねばならない。

本節では「Source Han Sans のサブセット・フォント (SourceHanSansHW-Regular.otf; ここには「令和合字」の縦横 2 グリフも含まれている)、Libertinus, UnifrakturMaguntia, PrincePlain.ttf」をウェブフォントへ変換⁴¹したものを論者運用のウェブサーバ内に用意し、当該フォントがシステムにインストールされていない環境でも令和合字、「ビャンビャン麺」文字、エスツェット大文字、変種ウムラウトを含むフラクトゥア、長い s、プリンスのラブ・シンボルが閲覧者のブラウザ上に表示されるような HTML サンプルファイルを作成した。

⁴⁰ <https://www.w3.org/TR/2018/REC-css-fonts-3-20180920/> (2020 年 9 月 18 日閲覧)

⁴¹ Source Han Sans と PrincePlain についてはフリーのフォント変換ソフト「武蔵システムの WOFF コンバータ」を使用した。<https://opentype.jp/woffconv.htm> (2020 年 9 月 18 日閲覧) Libertinus のウェブフォントは CUFON Fonts から入手できる。<https://www.cufonfonts.com/font/libertinus-sans> (2020 年 9 月 18 日閲覧) UnifrakturMaguntia についてはウェブ上のフリーのフォント変換サービス FONT SQUIRREL を利用した。<https://www.fontsquirrel.com/tools/webfont-generator> (2020 年 9 月 18 日)

[ウェブフォント：FU_box (ダウンロード用)]

06_webttypography.html

[ウェブフォント：Web (ウェブ閲覧用)]

[https://lgs.hum.fukuoka-u.ac.jp/~ynagata/2020/jinbun-ronso/
06_webttypography.html](https://lgs.hum.fukuoka-u.ac.jp/~ynagata/2020/jinbun-ronso/06_webttypography.html)

おわりに

ユニコードそして OpenType フォントは、現在なお日々目覚ましく発展し続けている。「コンピュータ・ディスプレイにおけるトーフの出現に終止符を打つ」といった壮大な目標を掲げて活動を続けるゲーグルのオープンソース・プロジェクト「Noto Fonts⁴²」は、コンピュータ上で文字そのもの、そして、正確な文字を使った多様で奥深い表現に拘りたい全ての者にとって、今後とも目が離せない対象の一つであるし、他者との一意なテキストデータ交換や精緻なテキストデータの運用が必要な者にとって、ユニコードを正しく理解することはますます重要になっていこう。例えば、信頼の置ける「全文検索」(4 バイト文字を含む多言語対応の有無、半角・全角・濁点の区別の有無、欧文大文字・小文字の区別の有無、等々) サービスといったようなアプリケーションソフトウェアの作成を考えるならば、データベースを作成する「前」に、文字コード及び照合順序 (Collation: 文字データの並び順に関する基準) の設定等を「予め決定しておく」必要があるのである。

かつて携帯電話端末で用いられていた日本発の「絵文字」は、今では Emoji として全世界に普及し、2010 年のバージョン 6.0 よりユニコードに取り入れられた。言い換えると、ユニコードに登録された Emoji の一つ一つは、今や一意のコードポイントを持つれっきとした「文字」となったのである。Emoji は、

⁴² <https://www.google.com/get/noto/> (2020 年 9 月 18 日閲覧)

しかも、1文字を2バイト以上で表す（コンピュータ上で取り扱うにはそれなりに複雑な）マルチバイト文字であり、漢字を含む日本語文字の殆どが3バイト（ただし第3及び第4水準漢字の多くは4バイト）に収まるのに対し、4バイトが基本で、さらに「字形選択子⁴³」(Variation Selector)や「絵文字列⁴⁴」(Emoji Sequence) という仕組みにより、「1文字」が原理上「何バイトにも成り得る」という仕様となっている。この先新たに登録されることになる Emoji「1文字」が、果たして何バイトまで拡張されるのか、想像もつかない。現時点においてさえ、4バイト文字をデータベースに格納するには注意が必要である⁴⁵。

本稿執筆途中の2020年8月26日、「来年2月開催のベルリン映画祭では最優秀『男／女』優賞を廃止し、新たに最優秀『主演／助演』賞を設ける」といったニュースが飛び込んできた⁴⁶。「ジェンダー」を意識する世の趨勢は、ユニコードの世界とも無縁ではない。例えば、Emoji「家族」(U+1F46A; FAMILY)は2018年までは「家族：男，女，少年」(U+1F468/U+1F469/U+1F466の3つのEmojiを「ゼロ幅接合子」(ZWJ; Zero Width Joiner)で1文字とした12バイトのEmoji; FAMILY: MAN, WOMAN, BOY)と同じであったが、2019年からは「三者から全て性徴を剥奪」した意匠（各プラットフォームによりデ

⁴³ 例えば、同じ向きを指す原型の「矢印」Emojiにこのセレクトを付加することでデザインの異なる「別の矢印」Emojiを表す方式。

⁴⁴ 囲み絵文字 (Emoji Combining Sequence), 皮膚色絵文字 (Emoji Modifier Sequence for Skin Tone), 国旗絵文字 (Emoji Flag Sequence), 家族絵文字・職業絵文字 (Emoji ZWJ Sequence), 等々。

⁴⁵ 現在世界で最も普及しているオープンソースのリレーショナルデータベース管理システム (RDBMS: Relational DataBase Management System) の一つであるMySQLでは、charset (文字集合およびエンコーディング) を utf8 で設定してデータベースを構築してしまうと、4バイト文字が正確に取り扱えなくなってしまう。例えば、1から3バイト目までが全く同じ「寿司Emoji」(F0 9F 8D A3)と「ビールEmoji」(F0 9F 8D BA)は、データベースの中では「同じ文字扱い」となる（いわゆる「寿司・ビール問題」）。4バイト文字を取り扱うには、予め charset を utf8mb4 (MultiByte4) としてデータベースを構築せねばならない。

⁴⁶ <https://www.rbb24.de/kultur/beitrag/2020/08/berlinale-2021-publikum-preise-neu-strukturiert.html> (2020年9月18日閲覧)

ザインは異なる)へと変更された。ユニコード・バージョン13のEmojiには、当然のように、「FAMILY: MAN, MAN, GIRL」, 「FAMILY: WOMAN, BOY, BOY」, 「KISS: WOMAN, WOMAN」, 「COUPLE WITH HEART: MAN, MAN」等々が登録されている。この先、バイオテクノロジーが一層の進展を遂げたとき、例えば「FAMILY: MAN, CAT, CHIMERA」のようなEmojiも誕生するのであろうか。

最後に、Emojiのようなマルチバイト文字を「テキスト」として本格的に取り扱うのであれば、「検索」の問題（どこまでが「1文字」なのか）は避けて通れない考察対象の一つである。この点については、稿を改めじつくりと論考してみたい⁴⁷。「皮膚色絵文字」を含むEmojiは「色刷り印刷ではない紙媒体」において「意図される文字の再現が不可能」となる、といったことも含め、「そもそも文字とは何なのか」といった問いに、今、改めて向き合わされているようにも感じる。

参考文献

- [1] Adobe: *How To Download The Pan-CJK Fonts From GitHub*.
<https://github.com/adobe-fonts/source-han-serif/raw/release/download-guide-source-han.pdf> (2020年9月18日閲覧) 以下、URL表記のあるものはいずれも同日に最終閲覧チェックをした。
- [2] Adobe: *Source Han Sans Version 2.001 ReadMe (Released 2019-04-09)*.
<https://github.com/adobe-fonts/source-han-sans/raw/release/SourceHanSansReadMe.pdf>

⁴⁷ 例えば、中世の欧文テキストに含まれる多種多様な文字および記号をフリーなOpenTypeフォントとして提供すべく開発が続けられているJunicond（バージニア大学のPeter S. Bakerによる）では、第15及び16面「私有面A及びB」に収録される文字の検索についても考究されている。レベルは異なるが、問題の本質は「大文字Sにマッピングされたラブ・シンボル」と同じところにある。<https://github.com/psb1558/Junicond-New> (2020年9月18日閲覧)

- [3] Adobe: *Syntax for OpenType features in CSS*.
<https://helpx.adobe.com/fonts/user-guide.html/fonts/using/opentype-syntax.ug.html>
- [4] adobe-fonts/source-han-sans (GitHub).
<https://github.com/adobe-fonts/source-han-sans>
- [5] adobe-fonts/source-han-serif (GitHub).
<https://github.com/adobe-fonts/source-han-serif>
- [6] alerque/libertinus (GitHub).
<https://github.com/alerque/libertinus>
- [7] CUFON Fonts: *Libertinus Sans Font Family*.
<https://www.cufonfonts.com/font/libertinus-sans>
- [8] Google Noto Fonts.
<https://www.google.com/get/noto/>
- [9] Deutsche Rechtschreibung: *Regeln und Wörterverzeichnis*. Aktualisierte Fassung des amtlichen Regelwerks entsprechend den Empfehlungen des Rats für deutsche Rechtschreibung 2016.
https://www.rechtschreibrat.com/DOX/rfdr_Regeln_2016_redigiert_2018.pdf
- [10] Rat für deutsche Rechtschreibung: *PRESSEMITTEILUNG: 29.6.2017*.
https://www.rechtschreibrat.com/DOX/rfdr_PM_2017-06-29_Aktualisierung_Regelwerk.pdf
- [11] psb1558/Junicode-New (GitHub).
<https://github.com/psb1558/Junicode-New>
- [12] Map of Lilliput and Blefuscu.
<https://www.bl.uk/collection-items/map-of-lilliput-and-blefuscu>
(2020年9月18日確認)
- [13] PrincePlaintf.
<https://www.dropbox.com/sh/jvj2ss4jfyvmvirl/AADXW3RWGwgYkjlmgSInvCRBa?dl=0>
- [14] Unicode: *Unicode 5.1.0. Released: 2008 April 4*.

- <http://www.unicode.org/versions/Unicode5.1.0/>
- [15] Unicode: *The Unicode Standard. Version 13.0 – Core Specification.*
<http://www.unicode.org/versions/Unicode13.0.0/UnicodeStandard-13.0.pdf>
- [16] Unicode: *Unicode 13.0 Character Code Charts.*
<https://www.unicode.org/charts/>
- [17] Unicode: *Latin Extended Additional.*
<https://www.unicode.org/charts/PDF/U1E00.pdf>
- [18] Unicode: *Letterlike Symbols.*
<https://unicode.org/charts/PDF/U2100.pdf>
- [19] Unicode: *Mathematical Alphanumeric Symbols.*
<https://unicode.org/charts/PDF/U1D400.pdf>
- [20] Rundfunk Berlin-Brandenburg: *Berlinale 2021 mit Publikum und ohne Geschlechtertrennung.*
<https://www.rbb24.de/kultur/beitrag/2020/08/berlinale-2021-publicum-preise-neu-strukturiert.html>
- [21] FONT SQUIRREL: *Webfont Generator.*
<https://www.fontsquirrel.com/tools/webfont-generator>
- [22] UnifrakturMaguntia: a Unicode Fraktur Font.
<http://unifraktur.sourceforge.net/maguntia.html>
- [23] Wehner, Mike: *We found Prince’s original Love Symbol font, and here’s how to use it.*
<https://www.dailydot.com/debug/prince-font-install-love-symbol/>
- [24] W3C: *CSS Fonts Module Level 3.* W3C Recommendation 20 September 2018.
<https://www.w3.org/TR/2018/REC-css-fonts-3-20180920/>
- [25] 永田善久 「No more Tofu!」 2018 年 5 月.
<https://lgs.hum.fukuoka-u.ac.jp/noto/>
- [26] 永田善久 「『令和』時代のユニコードと OpenType フォント」 2020 年 5 月.
https://lgs.hum.fukuoka-u.ac.jp/~ynagata/ict2b/unicode_reiwa.html
- [27] 武蔵システム WOFF コンバータ.

<https://opentype.jp/woffconv.htm>

- [28] *Das Lalebuch*. Nach dem Druck von 1597 mit den Abweichungen des Schiltbürgerbuchs von 1598 und zwölf Holzschnitten von 1680. Hrsg. von Stefan Ertz. Philipp Reclam Jun. Stuttgart, 1982.
- [29] *Duden Band 1. Die deutsche Rechtschreibung*. 22., völlig neu bearbeitete und erweiterte Auflage. Hrsg. von der Dudenredaktion. Auf der Grundlage der neuen amtlichen Rechtschreibregeln. Mannheim/Leipzig/Wien/Zürich 2001.
- [30] Voß, Herbert: *Einführung in LaTeX*. Unter Berücksichtigung von pdfLaTeX, XeLaTeX, LuaLaTeX. dante lehmanns media, Berlin 2012.
- [31] Voß, Herbert: *Mathematiksatz mit LaTeX*. Zweite, überarbeitete und erweiterte Auflage, dante lehmanns media, Berlin 2012.
- [32] Ken Coar et al./ 笹井訳『Apache クックブック』, オライリー・ジャパン, 2008 年.
- [33] Rasmus Lerdorf et al./ 高木訳『プログラミング PHP』, オライリー・ジャパン, 2007 年.
- [34] Eric A. Meyer/ クイープ訳『CSS 完全ガイド』, オライリー・ジャパン, 2006 年.
- [35] Sasha Pachev/ 菅野訳『詳解 MySQL』, オライリー・ジャパン, 2007 年.
- [36] 奥村晴彦・黒木裕介『改訂第 7 版 LaTeX2e 美文書作成入門』, 技術評論社, 2017 年.
- [37] スウィフト／平井訳『ガリヴァー旅行記』, 岩波文庫, 1985 年.
- [38] 宮沢賢治・草野心平編『宮沢賢治詩集』, 新潮文庫, 1979 年.