

メールから見えてくるコミュニケーションの本質

文字コードに思う

IIJ技術研究所 山本和彦

IIJ技術研究所でIPv6、セキュリティ、メールの研究に従事。おもな著書に「リスト遊び - Emacsで学ぶLispの世界 - (アスキー)などがある。趣味はジョギング、サッカー、吟醸酒、ワイン。最近フリーライミングに凝り始めている。

私は、Mewというフリーのメールリーダーを作り続けて、足掛け7年になる。MewのプラットフォームとなっているEmacsというエディターが、数多くの文字コードを扱う機能を提供しているおかげで、Mewでもさまざまな文字コードを使いメールをやり取りできる。このためか、Mewは日本に限らず世界各国で使われている。

昔から私は、多くの文字を持つ言語に慣れていたし、またJIS X 0208から派生したISO-2022-JP、EUC、シフトJISといった文字コードが混在する複雑な環境も理解していた。西洋の文字コードは、東洋の文字コードに比べて文字数が少ないためわかりやすい。実際、知りうる限りの文字コードをMewで

サポートし、経験を積んでいたのも、もう文字コードについて学ぶべきことはそんなに多くないと思っていた。

そんな考えを根底から覆すメールを、ドイツのMewユーザーから受け取った。アルファベットにアクセントの付くドイツ文字は、ISO-8859-1という文字コードで表現できる。ドイツ語を扱うためにはISO-8859-1をサポートしていれば十分なはずである。しかし、そのドイツ人はルーマニアに住んでいるドイツ人とメールのやり取りをする際に、ISO-8859-2も使いたいという。

たしかにドイツでは、ISO-8859-1が利用されている。一方、ルーマニアではルーマニア文字を表現するために、ISO-8859-2が使わ

れている。実はこの文字コードでもドイツ文字を表わすことが可能である。

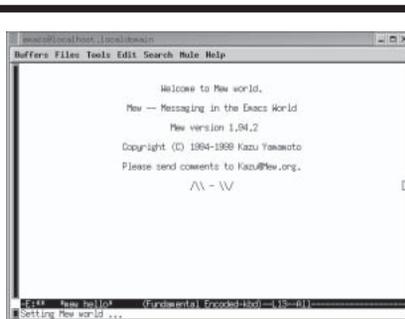
彼の要望はこうだ。ルーマニアに住んでいるドイツ人が、彼にメールを書くときには、何も考えずにISO-8859-2で送ってくる。したがって、そのメールに返答する際には、入力コードをISO-8859-1からISO-8859-2へ自動的に切り替えたい。

魚は空気に触れてはじめて水の世界を知るといふ。われわれは異文化に触れてはじめて、自分たちの文化を再認識する。私には最初、そのメールの内容がまったく理解できなかった。しかし、なんとか理解したそのときこそ、ヨーロッパでの文字コードの問題に実際に触れ、日本語の文字コードを再認識した瞬間

Mew

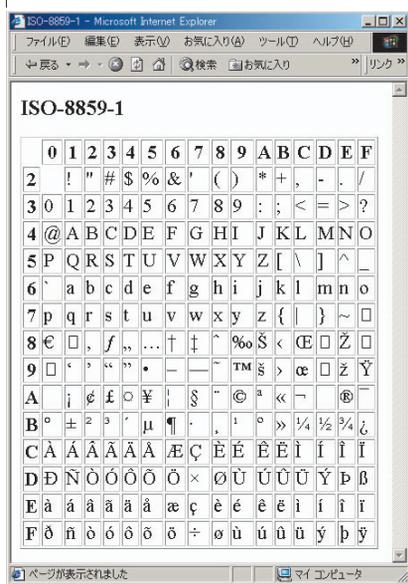
おもにUNIXで広く利用されているテキストエディター「GNU Emacs」の上で動作する、メールリーダーの名称。Emacsをベースに拡張したXEmacsやMule、Emacsをウィンドウズに移植したMeadowなどでも動作する。

 www.mew.org



JIS X 0208

1978年にJIS（日本工業規格）によって規格化された文字コード。このJIS X 0208をもとに、実際にコンピュータで文字を扱う際の文字コードであるシフトJISやEUC、ISO-2022-JPなどが生まれた。JIS X 0208は1983年と1997年に若干の改正が加えられ、現在はさらに多くの文字を追加した「JIS X 0213」が定められている。



ISO-8859-1 / ISO-8859-2

ISO-8859は、ヨーロッパ系の諸言語で使われる文字を扱うために作られた文字コード体系で、ISO-8859-1は西欧系の言語、ISO-8859-2は東欧系の言語で使われる文字をそれぞれ扱う。このほかにも、ロシア語を扱うISO-8859-5、ギリシャ語を扱うISO-8859-7など、複数のISO-8859が存在する。

0x20 ~ 0xFFまでの文字コード表をウェブブラウザで表示した例。ISO-8859-1 (左)とISO-8859-2 (右)では、ASCIIからの拡張部分(0x80 ~ 0xFF)が異なり、この部分が文字化けの原因になる。

Unicode

マイクロソフト、アップル、IBMなどのコンピュータ企業からなる「Unicode コンソーシアム」によって制定された文字コードの名称。当初は世界中のすべての文字を2バイトコード(最大で65,536字)として扱うことを目標としていたが、現在ではさらに多くの文字を扱うための拡張が検討されている。

www.unicode.org

だった。

「ヨーロッパの文字コードは文字数が少ないから簡単だ」と考えるのは、大きな誤りである。ヨーロッパでは、国と民族が入り乱れ、さまざまな言語が話されているという日本では考えられない状況が存在する。一方で、日本語は文字数が多いといっても、話されているのはおもに日本だけである。

それまでの自分がいかに偏見に満ちていたのか、私は身にしみて思い知らされた。私以外にも、「日本の技術者は文字コードに精通しており、欧米の技術者は文字コードに疎い」という片寄った考えに陥っている人がいると思う。その証拠に、「Unicodeは嫌いだ」という日本人は多い。

Unicodeを忌み嫌う理由の1つとして、Unicodeは中国の漢字と日本語の漢字を統合していることがよく挙げられる。違う言語の文字を同一と見なすなんて、文化の侵略にも似た野蛮な行為だ。「漢字のよくわからない欧米人に、漢字を取り扱う文字コードを設計させてはならない」というわけである。しかし、漢字の統合を提案したのは中国人だという事実はあまり知られていない。

翻って考えるに、日本人は西洋人の抱える問題を理解していると言えるだろうか?ヨーロッパにおいて共通の通貨としてユーロが

必要であるように、共通の文字コードも必要なのである。

日本人はこうも反論する。中国の漢字と日本語の漢字を同一視したら、たとえば日本語の日記のなかに中華料理のメニューが書けないではないかと。しかし、このような複数の言語の文字を同時に取り扱うことは、結果的にはUnicodeの目的ではない。先のドイツとルーマニアの例も、ドイツ語だけを問題にしている。

Unicodeは、たとえば「日本で使用している」という外部情報とともに使う。これによって、中国語の文字と日本語の文字を切り替えられる。もし、Unicodeの策定過程で、このことに対する共通の理解があったなら、JIS X 0208と互換性を保つ方式が採択されたのではないかと。たとえば、「日本で使用している」というモードでは、漢字の部分にJIS X 0208を(若干の変換を加えて)利用する方法である。

しかし、残念ながらそうはならなかった。現在のUnicodeはJIS X 0208と互換性がなく、JIS X 0208ゆかりの文字コードへ変換する場合は、大きな対応表を用いる他ない。

違う文字コードの文字を同一視するという行為が、それほど野蛮でないということも指摘しておこう。中国本土では、字体を簡略

化した「簡体字」が使われており、文字コードとしてはGBKなどが利用されている。一方、周辺地域の香港や台湾では「伝統字」が使用されており、それを表現する文字コードはBig5である。

中国本土でも周辺地域でも、話されているのが中国語であることに変わりはない。したがって、GBKのどの文字とBig5のどの文字とが同一なのかを考える必要がある。中国語を話す人たちにとって、Unicodeはこの問題を考えるよい機会だったのかもしれない。

もちろん、Unicodeは文字コードとしてできが悪い。間違いの本質は、世界で利用されている文字の数を低く見積もった点にある。もし今からやり直せるなら、4バイトの素直な体系を設計できるだろう。しかし、もう引き返せないところまで来てしまった。そろそろ日本人もUnicodeにまじめに向き合うべきではないかと私は考えている。

私はUnicodeの策定にかかわっていないので、かかわった人たちの努力を批判する気持はまったくない。ただ、文字コードをとおして思うのは、いかに異文化を理解するのが困難かということだ。コミュニケーションの本質は相手を理解し、そして自分を理解してもらうことだという。私にとって、Unicodeはこのよき教訓である。



[インターネットマガジン バックナンバーアーカイブ] ご利用上の注意

このPDFファイルは、株式会社インプレスR&D(株式会社インプレスから分割)が1994年～2006年まで発行した月刊誌『インターネットマガジン』の誌面をPDF化し、「インターネットマガジン バックナンバーアーカイブ」として以下のウェブサイト「All-in-One INTERNET magazine 2.0」で公開しているものです。

<http://i.impressRD.jp/bn>

このファイルをご利用いただくにあたり、下記の注意事項を必ずお読みください。

- 記載されている内容(技術解説、URL、団体・企業名、商品名、価格、プレゼント募集、アンケートなど)は発行当時のものです。
- 収録されている内容は著作権法上の保護を受けています。著作権はそれぞれの記事の著作者(執筆者、写真の撮影者、イラストの作成者、編集部など)が保持しています。
- 著作者から許諾が得られなかった著作物は収録されていない場合があります。
- このファイルやその内容を改変したり、商用を目的として再利用することはできません。あくまで個人や企業の非商用利用での閲覧、複製、送信に限られます。
- 収録されている内容を何らかの媒体に引用としてご利用する際は、出典として媒体名および月号、該当ページ番号、発行元(株式会社インプレス R&D)、コピーライトなどの情報をご明記ください。
- オリジナルの雑誌の発行時点では、株式会社インプレス R&D(当時は株式会社インプレス)と著作権者は内容が正確なものであるように最大限に努めましたが、すべての情報が完全に正確であることは保証できません。このファイルの内容に起因する直接のおよび間接的な損害に対して、一切の責任を負いません。お客様個人の責任においてご利用ください。

このファイルに関するお問い合わせ先

株式会社インプレスR&D

All-in-One INTERNET magazine 編集部

im-info@impress.co.jp