

■ 漢字(後編) (TBSラジオ「日本全国8時です」2004. 6. 24)

先週は、漢字は便利な文字ではあるが、種類が何万もあるために覚えるのが大変であり、それを解決しようと明治以来、漢字を廃止して仮名だけにしようとか、ローマ字にしようとか、極端な場合は、日本語そのものを廃止して英語やフランス語を国語にしようなど、様々な構想があったという話をさせていただきました。

そのような構想の影響を潜り抜けて、何とか1945文字の常用漢字と、287文字の人名漢字を中心にして、漢字を使う社会を維持しているのが日本の現状というわけです。

ところが、IT時代になり、別の問題が発生してきました。それは漢字をコンピュータに入力したり、インターネットで送信したりするときの問題です。

コンピュータやインターネットの仕組みに馴染みのない方のために、やや原理的なことから話をさせていただきますが、このような情報技術は数字や漢字をそのまま扱うことができず、「0」と「1」しか扱うことのできないのです。これを2進法といい、その単位を「ビット」といいますが、例えば、「6」という数字は「110」という数字に変更しないとコンピュータは「6」と理解してくれないのです。

同様に、アルファベットも仮名も漢字もすべて「0」と「1」に変更しておかないと、コンピュータは理解できないということになります。これを文字コードというのですが、例えば、アルファベットの大文字の「A」は、コンピュータの中では「0000010000000001」という16個も「0」と「1」が並んだ数字に置き換えています。

なぜ16個も「0」と「1」を並べなければいけないのかと疑問に思われるでしょうが、たくさんの文字を区別するためには、たくさんの数字が必要なのです。例えば「0」と「1」を2個しか並べないとすると「00」か「01」か「10」か「11」かの並べ方しかありませんから4種類の文字しか区別できません。

同様にして、「0」と「1」を4個並べれば16種類、8個並べれば256種類の文字が区別できることになります。コンピュータの分野では「0」と「1」を8個並べた単位を一区切りにして「バイト」と呼びますが、アルファベットだけであれば、大文字26文字と小文字26文字を合計しても52文字ですし、「0」から「9」までの10個の数字や、クエスチョンマークや括弧などの記号を加えても、100種類程度の文字が区別できれば十分ですから、1バイト、すなわち8個の「0」か「1」かの並べ方で表現できるということになります。

そこでアメリカやヨーロッパを中心にして発展してきたコンピュータの世界では、世界の様々な文字を含めても2バイト、すなわち16ビットもあれば十分だろうという考えで、文字コードが作られました。現在、世界規模で使用されているのは、1990年に、マイクロソフトやアップルコンピュータやIBMなどアメリカのコンピュータメーカーが中心になって作った「ユニコード」という文字コードです。これでも

6万5000文字程度が表現できますから、常用漢字程度なら問題ないだろうというわけです。

ところが、1994年に中国で発行された『中華字典』には8万5000字以上が採録されていますから、2バイトでは表現できないということになりますし、それ以外にも様々な問題があります。

まず、日本国内の問題では、戸籍で扱う文字の問題があります。例えば「吉田」の「吉」の字は、上の横棒が長い文字と下の横棒が長い文字がありますが、これはコンピュータの標準の文字コードでは上の棒が長い「吉」しか登録されていませんので、下の棒が長い「吉」は表現できないことになってしまいます。そうすると外字といって、各役所が文字を作ることになりますが、役所でそれぞれ作ったり、コンピュータメーカーがそれぞれ作ったりしますので、ある役所から別の役所に住民票の情報を送ると、「吉」の字が違う文字になってしまいます。

さらに大変なのは「渡辺」の「辺」の旧字体「邊」です。これは私も正確に書けないのですが、登録に来た人が間違えて書いてしまうと、戸籍はそのまま登録しますので、様々な「邊」ができてしまい、現在、戸籍で使われている「邊」は30種類近くあるといわれています。

ところが、世界全体から見ると、そのような細かいことには構ってられないということで、一種類にされてしまうということになります。

しかし、当人にしてみれば、先祖代々守ってきた名前を変えるわけにはいかないということにもなり、住民基本台帳ネットワークでは、これを解決するために、現在、どれだけ異なった文字があるかを調査している段階です。

同様に、古典の文書をコンピュータに入力しようとする、これもひとつの文字について様々な異字体があるのですが、勝手に標準の文字にしてしまうと、原典の意味がなくなってしまうことになりかねません。

さらに国際的な問題もあります。現在、漢字を使っている国は日本以外に、中国、台湾、韓国がありますが、同じ漢字でも、それぞれの国で微妙に形が違います。例えば、「骨」という漢字の上の部分は、日本では右側に曲がっていますが、中国では左側に曲がっています。漢字を使わない文化圏から見れば、同じにしか見えないので、一つの文字コードでいいのではないかということですが、それぞれの国民にとっては重要な問題です。

すなわち、漢字をコンピュータで扱ったり、インターネットで送ったりしようとすると、このような問題があるのです。これは一種の文化戦争でもあり、IT時代の根深い問題だと思えます。