

## 文字種の変換とフォントの罫（２）

Wordの[挿入]メニューから[記号と特殊文字]を選ぶと下のようなテーブルが表示されます。



赤枠内の文字は[種類]が「半角英字 1」で文字コードには 10 進数で 33～126 が割り当てられています。記号と特殊文字と呼んではいますが、キーボードから入力する通常の半角アルファベットと同じです。一方、青枠内の文字は[種類]が「半角英字 2」で文字コードには 10 進数で 161～255 が割り当てられています。0～32 まではタブや改行、スペースなど文字以外の諸々に使われているため、文字の割り当てはありません。この 0 から 255 まで 256 個が前回のコラムで書いた 1 バイトで表現可能な領域に相当します。



[記号と特殊文字]のテーブルで選択したときに Word の作業画面左下に数字が表示される文字があります。この数字が該当する文字の文字コードです（Word 2002 以降は[記号と特殊文字]でコード体系を選んで文字を選択すると一覧表の下にコードが表示されます）。左の例は A を選択した場合で、65 が A の文字コードに相当します。表示されるコードの頭が 0x で始まっている場合は 16 進数表記、単なる数字だけの場合は 10 進数表記です。参考までに、A の文字コード 65 を 16 進数で表す場合は 0x41 になります。

Windows の[スタート]メニューから[プログラム] [アクセサリ] [電卓]と選ぶと電卓が表示されます。この電卓を「関数電卓」にし、ラジオボタンで 10 進を選んだ状態で 65 を入力し、16 進のラジオボタンをクリックすると 16 進での値である 41 に変わります。これに 0x を付けた 0x41 が 16 進数でのコード表記です。Word で 16 進コードが表示される場合は上記と逆の方法で 10 進コードを計算することができます。普通の電卓を関数電卓にするには、電卓の[表示]メニュー（Windows NT の場合は[電卓の種類]）メニューから関数電卓を選択します。

[文字種の変更]で全角半角間に 1 : 1 の対応ができているのは、[記号と特殊文字]のテーブルで赤枠内にある文字です。この範囲についてはマクロなどを使わずに[文字種の変換]で全角 半角変換をしてもコードそのものが正しく変換されます。

一方、青枠内の文字はどうかというと、文字コードの範囲としては規格で定められた ASCII 半角文字の領域に相当します。ところが、日本語 Word に限って言えば、[記号と特殊文字]のテーブルから挿入される文字の中に規格の半角文字ではなく日本語の全角文字が混じっています。

これはおそらく次のような理由によるものと思われます。

青枠内の文字のなかには、ASCII 規格の英文半角文字コードをそのまま使うと日本語の半角カナと衝突するものがあります。半角英数字と半角カナのどちらも表示できる環境で、ひとつの文字コードに 2 種類の異なる文字が割り当てられているのはコンピュータにとっては非常に困ります。文字コードを見てもどちらの文字を表示すればよいのか判断できないからです。このため、英語版の Word を日本語版にローカライズするときに、半角カナと衝突する文字については該当する全角文字のコードで代用して衝突を回避したのではないかという気がします。裏付けを取ったわけではないので推測にすぎませんが、これ以外に有意な理由は思い当たりません。実際、Word 97/98 を使って[記号と特殊文字]で文字を選んでいくと、全角で挿入される文字だけ 10 進ではなく現在の Unicode に相当するコードが作業画面左下に表示されます。これでは混乱のもとだと思ったのか、Word 2000 では全角で挿入される文字に対して日本語 S-JIS コードに Unicode が並記されるようになり、Word 2002 以降はコード体系ごとに区別して[記号と特殊文字]でコードが表示されるようになっています。

さて、日本語版 Word で[記号と特殊文字]のテーブルから全角文字のコードが挿入されるのは次ページの表にピンク色で示した文字です。この中には[記号と特殊文字]で英文フォントを選択して挿入すると見た目が半角になるものがありますが、これはあくまでも見た目の問題であって本来の半角文字ではありません。フォントを変えれば元に戻ってしまいます。

ピンク色以外の文字は、英文フォントだけでなく日本語でも Microsoft 系のフォントなら表示は ASCII 規格の文字になります。これは日本語 Word 独自（なのか Microsoft 独自なのかは微妙なところ）の仕様で、これらの半角文字を Word でコピーして Unicode 未対応の他のアプリケーションにペーストすると、違う文字に置き換わってしまうことがよくあります。また、メモ帳などがそうなのですが、ペーストした時点では違う文字になっていなくても、データを保存してファイルを開き直すと文字が変わっていることもあります。

こうしたケースで、これらの文字がどのように置き換わるかを次ページに一覧で示します。もともと全角の文字については当然ながら同じ全角文字が入ります。コード 172 の文字は Unicode 対応アプリケーションにはペーストでき、Unicode 未対応のエディタなどにペーストしようとするとう消えてしまいます。純粋な英語環境であればこのようなことは起こり得ないのではないかと思います。日本語環境で使用する[記号と特殊文字]にはいろいろと注意が必要ですね。

文字	16進	10進		置換文字
¡	0xA1	161	inverted exclamation	!
¢	0xA2	162	cent sign	全
£	0xA3	163	pound sign	全
¤	0xA4	164	intl. currency sign	?
¥	0xA5	165	yen sign	¥
¦	0xA6	166	broken vertical bar	
§	0xA7	167	section sign	全
¨	0xA8	168	dieresis (umlaut)	全
©	0xA9	169	copyright sign	c
<sup>a</sup>	0xAA	170	feminine ordinal	a
«	0xAB	171	left double-angle quote	全
¬	0xAC	172	not sign	
–	0xAD	173	soft hyphen	-
®	0xAE	174	registered trademark sign	R
ˉ	0xAF	175	macron	全
°	0xB0	176	degree sign	全
±	0xB1	177	plus-minus sign	全
<sup>2</sup>	0xB2	178	superscript two	2
<sup>3</sup>	0xB3	179	superscript three	3
´	0xB4	180	acute accent	全
μ	0xB5	181	micro sign	全
¶	0xB6	182	paragraph sign	全
·	0xB7	183	middle dot	全
¸	0xB8	184	cedilla	全
<sup>1</sup>	0xB9	185	superscript one	1
<sup>o</sup>	0xBA	186	masculine ordinal	o
»	0xBB	187	right double-angle quote	全
<sup>1</sup> / <sub>4</sub>	0xBC	188	fraction one-fourth	?
<sup>1</sup> / <sub>2</sub>	0xBD	189	fraction one-half	?
<sup>3</sup> / <sub>4</sub>	0xBE	190	fraction three-fourths	?
¿	0xBF	191	inverted question mark	?
À	0xC0	192	A grave	A
Á	0xC1	193	A acute	A
Â	0xC2	194	A circumflex	A
Ã	0xC3	195	A tilde	A
Ä	0xC4	196	A dieresis	A
Å	0xC5	197	A ring	A
Æ	0xC6	198	AE ligature	A
Ç	0xC7	199	C cedilla	C
È	0xC8	200	E grave	E
É	0xC9	201	E acute	E
Ê	0xCA	202	E circumflex	E
Ë	0xCB	203	E dieresis	E
Ì	0xCC	204	I grave	I
Í	0xCD	205	I acute	I
Î	0xCE	206	I circumflex	I
Ï	0xCF	207	I dieresis	I
Ð	0xD0	208	N tilde	D

文字	16進	10進		置換文字
Ñ	0xD1	209	N tilde	N
Ò	0xD2	210	O grave	O
Ó	0xD3	211	O acute	O
Ô	0xD4	212	O circumflex	O
Õ	0xD5	213	O tilde	O
Ö	0xD6	214	O dieresis	O
×	0xD7	215	multiply sign	全
Ø	0xD8	216	O stroke	O
Ù	0xD9	217	U grave	U
Ú	0xDA	218	U acute	U
Û	0xDB	219	U circumflex	U
Ü	0xDC	220	U dieresis	U
Ý	0xDD	221	Y acute	Y
Ƨ	0xDE	222	capital thorn	T
ß	0xDF	223	sharp s (sz ligature)	s
à	0xE0	224	a grave	a
á	0xE1	225	a acute	a
â	0xE2	226	a circumflex	a
ã	0xE3	227	a tilde	a
ä	0xE4	228	a dieresis	a
å	0xE5	229	a ring	a
æ	0xE6	230	ae ligature	a
ç	0xE7	231	c cedilla	c
è	0xE8	232	e grave	e
é	0xE9	233	e acute	e
ê	0xEA	234	e circumflex	e
ë	0xEB	235	e dieresis	e
ì	0xEC	236	i grave	i
í	0xED	237	i acute	i
î	0xEE	238	i circumflex	i
ï	0xEF	239	i dieresis	i
ð	0xF0	240	small eth	d
ñ	0xF1	241	n tilde	n
ò	0xF2	242	o grave	o
ó	0xF3	243	o acute	o
ô	0xF4	244	o circumflex	o
õ	0xF5	245	o tilde	o
ö	0xF6	246	o dieresis	o
÷	0xF7	247	divide sign	全
ø	0xF8	248	o stroke	o
ù	0xF9	249	u grave	u
ú	0xFA	250	u acute	u
û	0xFB	251	u circumflex	u
ü	0xFC	252	u dieresis	u
ý	0xFD	253	y acute	y
þ	0xFE	254	small thorn	t
ÿ	0xFF	255	y dieresis	y

置換文字の「全」は元の全角文字がそのまま出ること示します。

最後にギリシャ文字について付記しておきます。コラム『文字種の変換とフォントの罨』の第 1 回目で、全角のギリシャ文字に対して[文字種の変換]をしてもフォントが変わるだけでコードは全角のままだということについて触れました。

では、[記号と特殊文字]でフォントに Symbol を選択した上で、表示された一覧の中から該当するものを挿入するとどうなるでしょうか。答えは、キーボードから入力した半角アルファベットを後から Symbol に変えた場合と全く同じ結果になります。文字コードには半角アルファベットのコードが使われますので、どちらの方法で入力してもフォントを変えればギリシャ文字ではなくなり、テキストファイルとして保存すれば通常のアルファベットに戻ってしまいます。

Symbol フォントは純粋な英語環境の Word にもありますから、作成したデータを使用する相手が Word を使っていることが分かっているのであれば、Symbol のまま入れておいても特に問題ないと思います（全角から文字種の変換で見た目だけ変えて入れるよりは何倍もよいです）。もし相手が Word 以外のアプリケーションを使っているのであれば、何らかの記号の組み合わせで代替して入れておくか、綴れるところは綴りで入れておくなどの対応をするとよいかと思います。

文字の扱いは Word に限って言えば Word 2002 あたりから安定してきているようですが、それでもバージョンによって少しずつ違いが見られます。かたや、他社アプリケーションで Unicode をどのくらい正しく解釈できるのかは、現時点ではまったく分かりません。このようなことを勘案すると、翻訳データの作成時には特殊文字の扱いに対するクライアントの意向を事前に聞くようにし、よく分からない場合はキーボードから入力できる文字（+フォントの指定）の範囲でおさめておくのが無難でしょう。