

資 料

情報交換用符号の拡張と登録†

都 丸 敬 介** 島 谷 和 典†††

Abstract

In order to make use additional characters other than which assigned on a standard 7-bit or 8-bit code table, code extension techniques have been studied at ISO/TC 97 /SC 2 for several years, and a Draft International Standard (DIS) was published.

This paper presents basic rules of the code extension techniques and procedures for international registration of characters and related escape sequences.

1. ま え が き

情報交換用標準符号の規定としては、ISO/R 646-1967 (6単位符号および7単位符号)、JIS C 6220-1969 (7単位符号および8単位符号)等があり、各種キャラクタとそれに対応したビット組合せ、すなわちキャラクタ符号が規定されている。しかし、多くのシステムあるいは装置では、標準化されていないキャラクタが使用されている。また、ある種のシステムでは1つの7単位あるいは8単位符号表に収容しきれないほど多種類のキャラクタが使用される。このような場合に標準的な符号表に含まれていない任意のキャラクタを使用するための技術を符号の拡張法と呼ぶ。異なったシステム間で誤りのない情報交換を行なうためには、キャラクタ符号の標準化のみでなく、拡張法の標準化も必要である。更に標準化されていない符号表の一元的な登録・管理が必要である。これらの課題は、ISO/TC 97/SC 2††††で取上げられ、1971年以来研究が続けられている¹⁾。そして、1971年10月のSC 2第8回会議において、国際標準規格案の原案が作成された²⁾。その際の決議事項の1つとして、各国の委員は、各国内でこの技術の周知を計ることが要請されているので、本誌の紙面を借りて、符号拡張技術の要点と追加符号の国際的な登録法について述べる。

2. 符号表の構成

符号拡張法を理解するために、先ず7単位符号表および8単位符号表の構成ならびにキャラクタ等の標記法を規定する。図1および図2に7単位符号表および8単位符号表の基本的な構成を示す。符号表上のキャラクタ位置やキャラクタの略号は以下のように標記する。

7ビット組合せのビット:

$b_7 \ b_6 \ b_5 \ b_4 \ b_3 \ b_2 \ b_1$

8ビット組合せのビット:

$a_8 \ a_7 \ a_6 \ a_5 \ a_4 \ a_3 \ a_2 \ a_1$

列および行に対するビットの重み:

$2^3 \ 2^2 \ 2^1 \ 2^0 \ 2^3 \ 2^2 \ 2^1 \ 2^0$

列 行

ビット組合せは符号表における位置の列および行の番号で表わされる。列番号はビット a_8-a_5 (あるいは

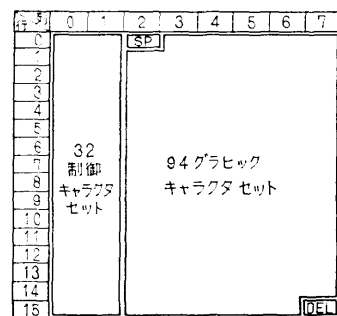


図1 7単位符号表の構成

Fig. 1 Structure of a 7-bit code table

† Code extension techniques and registration procedures, by Keisuke Tomaru (Nippon Telegraph and Telephone Public Corporation Yokosuka Electrical Communication Laboratory) and Kazunori Shimaya (Nippon Electric Co., Ltd., EDP Group Planning Division)

†† 日本電信電話公社 横須賀電気通信研究所

††† 日本電気株式会社 情報処理計画本部

†††† 国際標準化機構, 第97技術委員会第2小委員会

b_7-b_5) に対応する 10 進数であり、行番号はビット a_4-a_1 (あるいは b_4-b_1) に対応する 10 進数である。

10 進表示をする場合、8 ビットの列番号は 00 から 09 と頭にゼロをつける。たとえば間隔 (space) キャラクタの位置は 7 ビット符号表では 2/0 で、8 ビット符号表における同じキャラクタの位置は 02/0 である。

SO, ESC のようなキャラクタニモニックと、0/5, 1/7 のような列/行番号は、1 つのビット組合せのみに対応することを強調するために下線を引いて示す。

図 1 に示したように、7 ビット符号表は、8 列×16 行の 128 個の符号位置を持ち、以下のように制御キャラクタとグラフィックキャラクタに対する領域に群分けされる。

- (1) 列 0 と 1 に割当てた 32 種類の制御キャラクタセットに対する領域。
- (2) 位置 2/0 の間隔キャラクタ。これは制御キャラクタあるいは印刷されないグラフィックキャラクタのどちらとも見なすことができる。
- (3) 列 2 から 7 に割当てた 94 種類のグラフィックキャラクタセットに対する領域。
- (4) 位置 7/15 のまっ消 (Delete) キャラクタ。

8 ビット符号表は図 2 に示したように、列番号 00 から 15 と行番号 0 から 15 の 16×16 の配列により 256 の符号位置を持つ。

この配列の列 00 から 07 の 128 のキャラクタ位置は 7 ビットセットのキャラクタと 1 対 1 に対応する。列 08 と 09 は制御キャラクタ用であり、列 10 から 15 はグラフィックキャラクタ用である。ただし、位置 10/0 と 15/15 は除く。8 ビット符号の列 08 と 09 の制御キャラクタは伝送制御キャラクタを含まない。

列	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
0			SP													
1																
2																
3																
4																
5																
6																
7																
8																
9																
10																
11																
12																
13																
14																
15																

図 2 8 単位符号表の構成

Fig. 2 Structure of an 8-bit code table

3. 7 単位系における符号の拡張

3.1 拡張の単位

ISO/TC 97/SC 2 では、まず 7 単位系における拡張法を検討し、更にこれを 8 単位系に適用することを検討した。こうした経過から、7 単位系における拡張法が基礎になっている。符号の拡張とは、標準符号表に含まれているキャラクタを他のキャラクタに置換えることによってキャラクタの数を追加することである。追加キャラクタは、下記のように単独で追加する場合と、複数個のキャラクタの集合、すなわちキャラクタセットとして追加する場合がある。

- (1) 1 つのキャラクタを単独で追加する。
- (2) 32 個の制御キャラクタの集合を追加する。
- (3) 94 個のグラフィックキャラクタの集合を追加する。
- (4) 1 つのキャラクタが 1 バイト以上で表現される 94 個以上のグラフィックキャラクタの集合を追加する。

実用上はこれらを組合せて使用することが多い。

複数のキャラクタの集合を拡張単位とする場合の概念を明らかにするために、以下の 4 種類のキャラクタセットを定義する。

- C 0 セット: 32 制御キャラクタセット。
- C 1 セット: 32 制御キャラクタセット (追加セット)。
- G 0 セット: 94 グラフィックキャラクタセット (複数バイトセットも G 0 セットとして作用する)。
- G 0 セット: 94 グラフィックキャラクタセット (追加セット)。

これらの集合が符号表上のどの領域に対応し、どのように使われるかということについて以下の項で述べる。各集合の内容は、システムの要求に応じて決められるが、情報交換の目的を遂行するために以下の事項を守る必要がある。

- (1) 10 個の伝送制御キャラクタ、NUL, SO, SI, CAN, SUB, ESC, SP および DEL は、その意味および符号表上の位置を変えない。
- (2) ISO/R 646 のグラフィックキャラクタは他の位置に移してはならない。

- (3) キャラクタセットの全ての位置にキャラクタを割当てて必要はない。
- (4) 1つのキャラクタセットの中に他のキャラクタセット、たとえば ISO/R 646 で使用しているキャラクタがあっても差支えない。

3.2 符号拡張のためのキャラクタ

符号拡張の動作を実行するために、

ESC: 拡張 (Escape),

SI: シフトイン (Shift-In),

SO: シフトアウト (Shift-Out),

DLE: 伝送制御拡張 (Data Link Escape)

の4種類のキャラクタがあるが、DLE は伝送制御キャラクタの拡張のみに使用するものであり、ここで目的としている符号の拡張には使用しない。

後述するように、SI と SO は単独で使用するが、ESC は、ESC で始まる2つあるいはそれ以上の7ビット組合せからなるエスケープシーケンスの形で使用する。

3.3 指示, 呼出し, 表現

符号を拡張するために、

指示する (to designate),

呼出す (to invoke),

表現する (to represent)

と呼ぶ3種類の基本的な動作がある。

これらの動作をグラフィックキャラクタを例にとって図示したのが図3である。多数のグラフィックキャラクタセットの中から、G 0およびG 1として使用するものを選び出す操作のことを指示するという。

指示した2つのグラフィックキャラクタセットのうちのどちらを使用するかということを指定する操作を呼出すという。

指示され、呼出されたキャラクタセットにおけるあるビットパターンを、そのビットパターンに対応したキャラクタの意味に使用することを表現するという。

G 0 セットを呼出すには SI を使用し、G 1 セットを呼出すには SO を使用するが、グラフィックキャラクタセットの指示、制御キャラクタセットの指示と呼出し、および追加制御キャラクタの表現には、特定のエスケープシーケンスを使用する。

複数バイトグラフィックキャラクタセットはG 0の位置に指示される。追加グラフィックキャラクタセットを1組しか使用しない場合は、情報交換当事者間の同意があれば、エスケープシーケンスで指示しなくても、G 1の位置に指示されたものとして扱う。この場合、

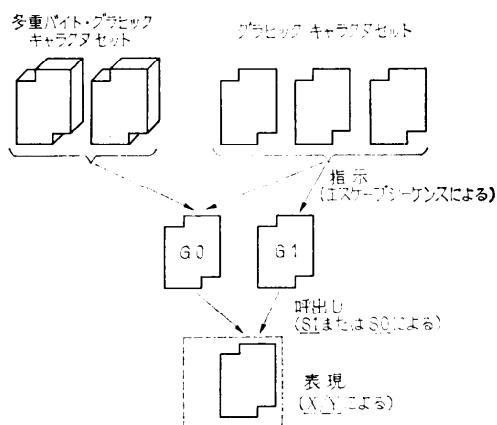


図3 符号拡張の基本動作 (グラフィックキャラクタの追加法)

Fig. 3 Basic operation of code extension techniques

G 0 の位置には標準のグラフィックキャラクタセット (たとえば国際的情報交換では ISO/R 646 で規定されたもの) が指示されたものとして扱う。

位置 2/0 の SP と 7/15 の DEL は、符号拡張法の影響を全く受けることがなく、不変である。

4. エスケープシーケンス

4.1 エスケープシーケンスの使用法

エスケープシーケンスの使用法には次のようなものがある。

- (1) グラフィックキャラクタセットの指示。
- (2) 制御キャラクタセットの指示と呼出し。
- (3) G 1 セットとして呼出された制御キャラクタの表現。
- (4) 単独で扱われる追加制御キャラクタの表現。
- (5) 情報交換の環境に7ビット系と8ビット系が混在している場合、それぞれの系でどのような符号拡張法を用いているかということのアナウンス (アナウンスシーケンス)。

4.2 エスケープシーケンスの構造

1つのエスケープシーケンスは、2つあるいはそれ以上の7ビット組合せからなる。最初は常に ESC のビット組合せであり、最後は常に最終キャラクタ (Final Character) である。エスケープシーケンスは、中間キャラクタ (Intermediate Character) を表わすいくつかの7ビット組合せを含むことがある。

エスケープシーケンスの意味は、中間キャラクタ (それがあつた場合) を表わす7ビット組合せと最終キ

キャラクタを表わす7ビット組合せで決まる。エスケープシーケンスの意味は、そのビット組合せだけで決り、個々のビット組合せに割当てられた意味には影響されない。中間キャラクタは、7ビット符号表の列2の16個のキャラクタである。これらの16個の中間キャラクタのうちの任意の1つを記号(I)で表わす。

最終キャラクタは、7ビット符号表の列3から7のうち、位置7/15を除く79個のキャラクタである。これらの79個の最終キャラクタのうちの任意の1つを記号(F)で表わす。列0と1および位置7/15の33個の制御キャラクタはエスケープシーケンスにおける使用が禁止されており、エスケープシーケンスを構成する中間キャラクタあるいは最終キャラクタのいずれにも使ってはならない。これらの禁止されたキャラクタが誤ってエスケープシーケンスに入っているときは、それを検出して誤りとして処理する手段を装置内に用意する必要がある。エスケープシーケンスは、それを構成するビット組合せの数により、

2キャラクタエスケープシーケンス: ESC (F),

3キャラクタエスケープシーケンス: ESC (I) (F),

4あるいはそれ以上のキャラクタのエスケープシーケンス: ESC (I) (I)……(I) (F)
等に分類される。

2キャラクタエスケープシーケンスは、追加制御キャラクタを表現するのに使用する。最終キャラクタ(F)は列3~7(7/15を除く)の79個のビット組合せである。(F)は表1のようにF_p, F_eおよびF_sの3種類に分類される。3キャラクタエスケープシーケンスは、単独追加制御キャラクタの表現、グラフィックキャラクタセットの指示、制御キャラクタセットの指示と呼出し等に使用する。表1のように、中間キャラクタ(I)は列2の16個のビット組合せであり、最終キャラクタ(F)は列3~7(7/15を除く)の79個のビ

ット組合せである。(F)は(F_p)と(F_i)の2種類に分類される。F_s, F_e, F_pおよびF_iの使い分けについては後で述べる。

4あるいはそれ以上のキャラクタのエスケープシーケンスの具体的な使用法については、未だ検討されていない。

4.3 グラフィックキャラクタセットの指示に使用するエスケープシーケンス

ESC 2/8 (F) と ESC 2/12 (F) は、G 0 セットとして使用する 94 グラフィックキャラクタセットを指示する。指示されたセットは SI で呼出される。ESC 2/9 (F) と ESC 2/13 (F) は G 1 セットとして使用する 94 グラフィックキャラクタセットを指示する。指示されたセットは SO で呼出される。ESC 2/4 (F) は、列2から7のうち、位置2/0と7/15以外のビット組合せの2つあるいはそれ以上のバイトで表わされるグラフィックキャラクタセットを指示する。指示されたセットは SI で呼出される。従って、これは G 0 セットと見なす。1つの多重バイトグラフィックキャラクタセットの中では、各グラフィックキャラクタは、同数のバイトで表わす。

ESC (I) (F_i) で指示するグラフィックキャラクタセットは、標準化のために登録されたものである。ESC (I) (F_p) で指示するグラフィックキャラクタセットは、システム毎に個別に規定するものであり、その意味は、情報交換当事者間で事前に同意しておく必要がある。

4.4 制御キャラクタセットの指示と呼出しに使用するエスケープシーケンス

ESC 2/1 (F) は列0と1のビット組合せで表わす32制御キャラクタセットをC 0 セットとして指示し、呼出す。

ESC 2/2 (F) は C 0 セットに影響を与えることなしに32制御キャラクタセットをC 1 セットとして指示し、呼出す。

C 1 セットには伝送制御キャラクタを含まない。

4.5 C 1 セットとして呼出された制御キャラクタの表現に使用するエスケープシーケンス

C 1 セットの32個のキャラクタの中の特定のキャラクタの表現には ESC (F_e) を使用する。C 1 セットとして、1種類の制御キャラクタセットしか必要としないときは、情報交換当事者間の同意があれば、指示および呼出しを省略して、ESC (F_s) で直接表現してよい。

エスケープシーケンス	分 類	列	キャラクタ数
2 キ ャ ラ ク タ	F _p	3	16
	F _e	4-5	32
	F _s	6-7 (7/15を除く)	31
3キャラクタ以上	I	2	16
	F _p	3	16
	F _i	4-7 (7/15を除く)	63

第1表 中間キャラクタと最終キャラクタの割当
Table 1 Assignment of intermediate characters
and final characters.

4.6 単独追加制御キャラクタの表現に使用するエスケープシーケンス

単独追加制御キャラクタの表現には、ESC (F_s), ESC (F_p) あるいは ESC 2/3 (F) を使用する。ESC (F_s) で表現する 31 個の制御キャラクタは標準化したものである。ESC (F_p) で表現する 16 個の制御キャラクタは標準化せず、情報交換当事者間で個別に決めてよい。ESC 2/3 (F) は、ESC (F_s) あるいは ESC (F_p) では必要な個数の追加制御キャラクタが得られない場合などに使用できる。

4.7 アナウンスシーケンス

ESC 2/0 (F) はその後に続くデータに適用されている拡張法のアナウンスに使用する。アナウンスの詳細は 6 節で述べる。

4.8 エスケープシーケンスの省略

情報交換当時者同士が唯一の G 0 セット、唯一の G 1 セット、唯一の G 0 セットおよび唯一の C 1 セット (あるいはこれらと同数の任意のセットが用いられる場合) のみを使用することに同意した場合は、これらを指示し、呼出すためのエスケープシーケンスを省略することに同意してもよい。しかし、このような同意は、データを次々に交換する機能を減少させることになるので注意を要する。

4.9 7 単位系におけるエスケープシーケンスのまとめ

以上述べた各種のエスケープシーケンスとその作用を 7 単位系の符号拡張に関して一括して図示すると図 4 のようになる。

5. 8 単位系における符号の拡張

8 単位符号表は図 2 のように、

- (1) C 0 セットとして扱う 32 個の制御キャラクタ (列 00, 01),
- (2) C 1 セットとして扱う 32 個の制御キャラクタ (列 08, 09),
- (3) G 0 セットとして扱う 96 個のグラフィックキャラクタ (列 02~07, 但し 02/0 と 07/15 を除く),
- (4) G 1 セットとして扱う 96 個のグラフィックキャラクタ (列 10~15, 但し 10/0 と 15/15 を除く),
- (5) その他 (02/0, 07/15, 10/0 および 15/15) の各領域に分けられる。

各種のキャラクタセットの指示あるいは呼出しは、

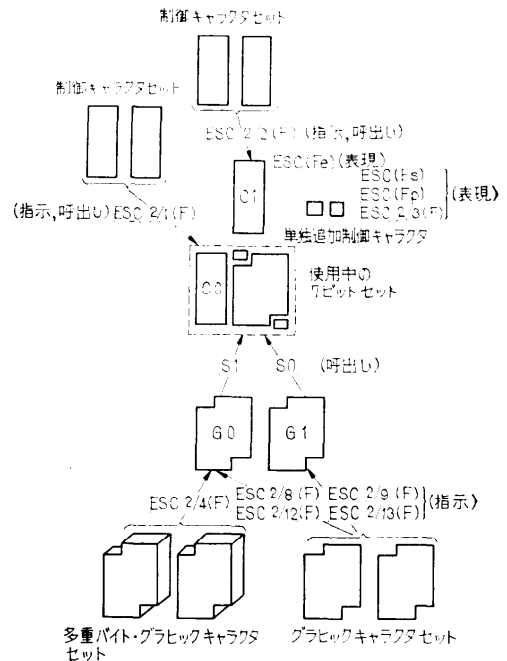


図 4 7 単位符号の拡張

Fig. 4 Extension of 7-bit code

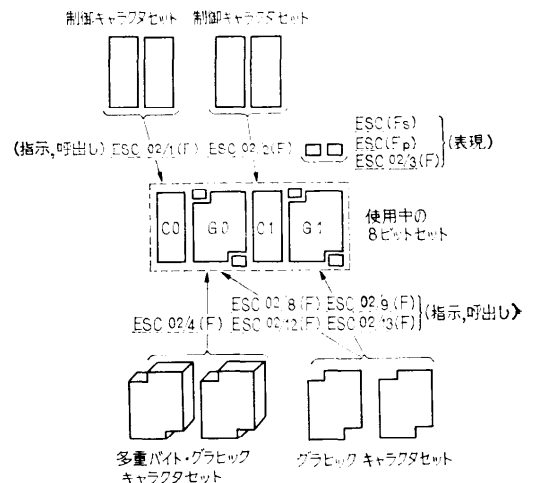


図 5 8 単位符号の拡張

Fig. 5 Extension of 8-bit code

7 単位系の場合と同様に、図 5 に示したようにエスケープシーケンスによって行なう。しかし、下記の点で 7 単位系の場合と異なる。

- (1) G 0 セットと G 1 セットの呼出しは指示と同時に行なわれる。したがって、SI と SO を使

用する必要はない。

- (2) C1 セットの中の任意のキャラクタは列 08 あるいは 09 の 1 つの 8 ビット組合せで表現されるが、7 単位系の場合のように ESC (F_e) を用いて表現することもできる。

- (3) 列 08~15 の全てのキャラクタをエスケープシーケンスに使用してはならない。これらのキャラクタがエスケープシーケンス中に現れた時は誤りとして処理する。

4 あるいはそれ以上のキャラクタのエスケープシーケンスも 7 単位系の場合と同様に解釈するが、具体的な使用法は未だ規定されていない。

6. 拡張法のアナウンス

8 単位媒体上などのような 8 単位環境では、すべてのキャラクタを 8 単位系で表わすことが望ましいが、7 単位環境と 8 単位環境との間の情報交換の容易さを重視して、8 単位系環境でも 7 単位系と同様に各キャラクタを表わすことがある。この場合の符号拡張法は、7 単位系における符号拡張法をそのまま使用する。

こうした種々の場合に、異システム間の情報交換を間違いなく行なうために、以下に続くデータのグラフィックキャラクタ集合が、7 単位環境および 8 単位環境のそれぞれにおいてどのように表わされているかということを経初めに通知するために、中間キャラクタが 2/0 の 3 キャラクタエスケープシーケンスを使用する。これをアナウンスシーケンスと呼び、その分類は最終キャラクタで表わす。最終キャラクタとそれが表わす状態は図 6 の通りである。ESC 2/0 4/1 の場合はグラフィックキャラクタは G0 のみであり、それ以外では G0 と G1 が使用できる。

5 節で述べたように、8 単位環境における C1 の位置は、8 単位符号表の列 08 と 09 であるが、その中のキャラクタの表現に ESC (F_e) を使用できるので、実質的には、7 単位系における C1 の位置にあるものとみなしてよい。したがって、アナウンスシーケンスでは、グラフィックキャラクタ集合の位置についてのみ考慮していると考えることができる。

情報交換の当事者間の同意があれば、アナウンスシーケンスを省略してよい。

7. 7 単位符号と 8 単位符号の変換

アナウンスシーケンス ESC 2/0 4/2 で示される場

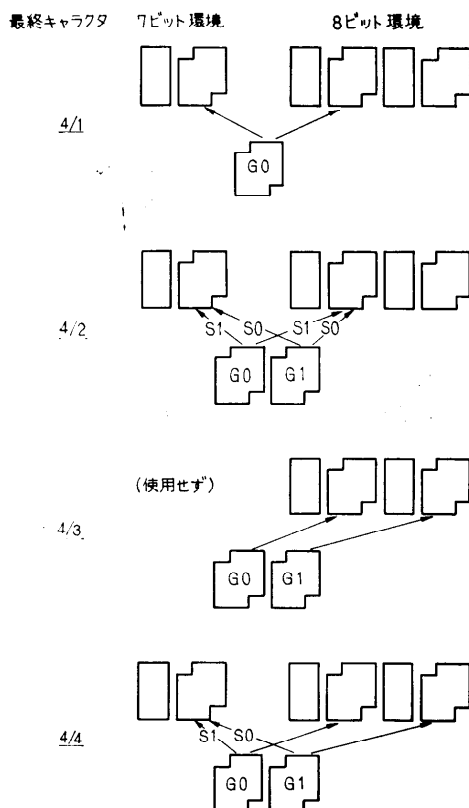


図 6 アナウンスシーケンスの意味

Fig. 6 Meaning of announcer sequences

合には、7 単位符号の $b_7 \dots b_1$ はそれぞれ 8 単位符号の $a_7 \dots a_1$ に対応し、 $a_8=0$ としておけばよいので、環境が 7 から 8 へ、あるいは 8 から 7 に変わっても、 $a_8=0$ をつけるか、あるいは落すか以外の符号変換は必要でない。しかし、ESC 2/0 4/4 で示される場合には、環境の変化に伴う符号の変換が必要である。この変換手順は IISO/TC 97/SC 2 で検討されているが、紙面の都合で省略する。

8. エスケープシーケンスの登録

これまで述べた各種のエスケープシーケンスには、アナウンスシーケンス以外は未だ固定的な意味が割当てられていない。国際的な情報交換のためには、ESC (F_p) や ESC (I) (E_p) のように個別に指定することを許されているもの以外のすべてのエスケープシーケンスと、これらのエスケープシーケンスで指示、呼出し、あるいは表現されるキャラクタあるいはキャラクタセ

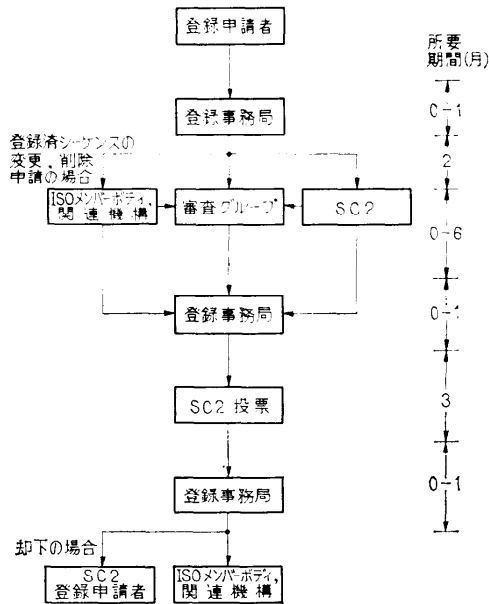


図7 エスケープシーケンスの登録、審査の手続き
Fig. 7 Procedures for registration of escape sequences

ットの一元的な登録、管理が必要である。

こうした登録および管理の主管は ISO であるが、実行は ISO/TC 97/SC 2 の事務局である AFNOR (フランス標準化協会: Association Française de Normalisation) が担当することになっており、登録手続、申請された書類の審査法等が SC 2 で検討されている^{3),4)}。この登録法に関する文書は、国際規格 (International Standard) として制定するための手続が取られており、すでにいくつかの国から登録申請が行なわれている。登録申請および審査は図7の手順で行なわれる。申請書の書式も規定されている。

登録申請ができる者は、下記の範囲に限定されている。

- (1) ISOの技術委員会(TC)あるいは小委員会(SC)。
- (2) 情報処理における符号化に関連した SC 2 内の任意のグループで、符号拡張あるいはエスケープシーケンスの使用のために指定されたものの。
- (3) ISO の任意のメンバーボディ。
- (4) ISO あるいはその技術委員会または小委員会と関連ある任意の国際機構。

登録事務局の AFNOR は、登録申請書を受取ると、1ヵ月以内にその写しを ISO のメンバーボディ、関

連機構および審査グループに配布する。

審査グループは3名で構成され、SC 2 の会議で指名される。したがって、任期は次の SC 2 会議までであるが、再任は許される。

登録申請書の実質的な審査は、この審査グループが担当し、特に次の点について検討する。

- (1) 登録済のものとの重複。
- (2) 符号化に関する ISO 文書との合致性。
- (3) 申請されたキャラクタセットの中に付加的に使用される代替を含まないこと。もしも付加的な代替キャラクタが必要な場合は、別個に申請する。

上記検討結果は次のように分類して報告される。

- (1) 申請をそのまま受理する。
- (2) 申請を登録済のものと同じ物として受理し申請者の名称のみを登録簿の申請者欄に追記する。
- (3) 修正を要求する。
- (4) 登録を拒絶する。この場合は理由を明らかにする必要がある。

審査グループが検討した結果、申請を受理してもよいという報告があると、登録の可否が SC 2 のメンバーによる郵便投票にかけられる。投票依頼から3ヵ月以内に2票以上の SC 2 の P メンバーの反対がなければ、その申請は登録される。しかし2票以上の SC 2 の P メンバーの反対があった時は、申請者は SC 2 内の郵便投票を請求できる。その結果、賛成多数であれば申請は登録される。

新しい申請が登録されると、事務局は、申請者、ISO の全メンバーおよび ISO 関連機構にその事を通知する。登録拒絶の場合は、事務局は、申請者、SC 2 のメンバーおよび SC 2 関連機構にその事を通知する。

9. エスケープシーケンスの登録の現状

前節で述べた手順に従って、すでに米国、英国等から国内規格を中心に登録申請書が提出されている。わが国でも SC 2 国内委員会において、“JIS C 6220-1969 情報交換用符号”で規定されている符号表の登録申請が検討されている。また種々の装置の制御キャラクタに対するエスケープシーケンスを標準化するための検討が、SC 2 において精力的に進められている。この検討は、印刷機能を持つ装置とディスプレイ機能を持つ装置を中心に、制御機能をリストアップし、それを標準化という面から見て優先順位をつけて分類し、そ

それぞれの制御キャラクタを使用するためのエスケープシーケンスを割当てるという順序で進められている。

10. む す び

符号拡張技術は 1967 年の SC 2 第 4 回会議で検討開始が決定されてから 4 年目に国際規格案の原案がまとまったが、これで検討が終了したわけではない。

たとえば、グラフィックキャラクタセットを G 0 と G 1 の 2 種類だけではなく、G 0, G 1, G 2, G 3 の 4 種類とし、これらを出すのに SI と SO の他に ESI と ESO を使う技術の検討などが残されている。

各装置固有の追加制御キャラクタを表現するエスケープシーケンスの使用や登録は、国際的な情報交換の

実施のために大切なことである。こうしたことから、本文が関係各位の参考になれば幸いである。

参 考 文 献

- 1) 情報処理学会：電子計算機の国際標準化。
- 2) ISO/TC 97/SC 2 N 562, Code extension techniques for use with the 7-bit coded character set of ISO/R 646.
- 3) ISO/S 2375, Procedure for registration of escape sequences.
- 4) ISO/TC 97/SC 2 N 580, Procedure within SC 2 for consideration of proposals of escape sequences to be registered according to ISO/S 2375.

(昭和 47 年 7 月 18 日受付)