

事務文書体系の基盤符号系の統一の提案

若鳥陸夫

日本ユニバック株式会社

国際標準化機構で国際規格制定の最終段階にある事務文書体系(ODA)の広範囲の応用仕様(application profile)において、漢字符号系の指示・呼び出し方式の統一が必要である。その応用仕様の対象はG4ファクシミリ(ミクスド・モード)、日本語ワードプロセッサ通信、パーソナルコンピュータ通信、電子郵便系及びデータベース系などに及ぶ。これらの異系統・異媒体での円滑な情報交換を行う統一的な漢字符号系として、JIS X0202に沿って、JIS X0208 情報交換用漢字符号系が拡張されることを前提として、「漢字集合を"G1"に指示し、符号拡張アナウンスにより、G0・G1をロックシフトなしで使用する」提案を行う。この使用方式によれば交換データ量を少なくできる。

A proposal on code extension of Japanese Kanji character.

- minimized data length for communications -

Rikuo WAKATORI

Nippon Univac Kaisha Ltd.

2-17-51, Akasaka, Minato-ku, Tokyo, 107, JAPAN

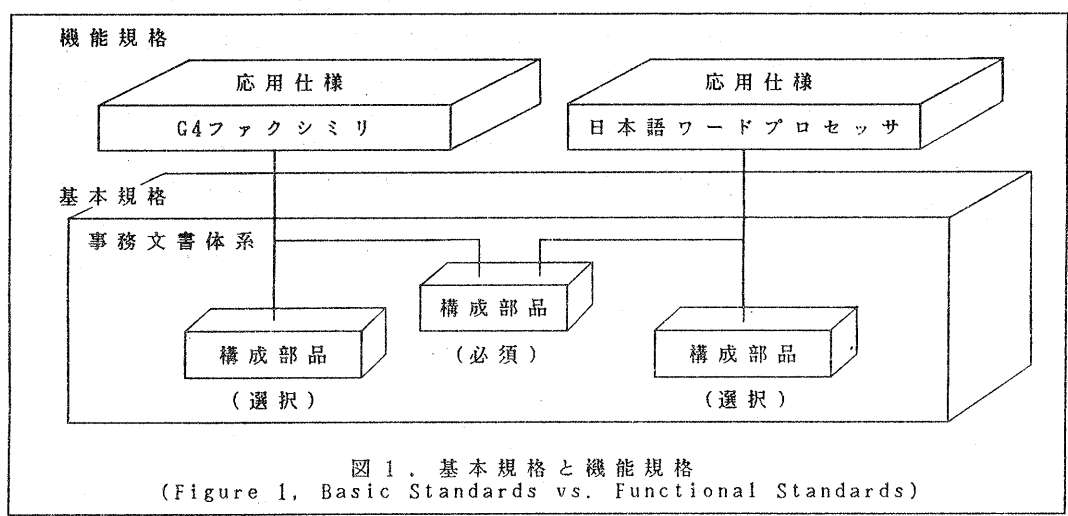
This paper proposes an effective usage of the code extension which will be applied to the Office Document Architecture(ODA).

Alphabet-Numeric code(JIS X0201-76) is designated to G0[Left] and Kanji code(JIS X0208) is designated to G1[Right]. And further more, code extension announcer specifies the usage of G0 and G1 without Locking-shift function. This method eliminates any escape sequences between alphabet/numeric code and two bytes Kanji code in data stream during information interchange. Same scheme may be applied to other multi-bytes code system, such as Chinese and others. (Written in Japanese.)

1 はじめに

開放型システム間相互接続(OSI)の方向に沿って、異系統・異媒体間の情報の再現性について考える。

国際標準化機構(ISO)で、国際規格化の最終段階にある事務文書体系(ODA)[文献1~14]は、開放型システム間相互接続及び在来形通信での事務文書の交換のための基本規格(basic standards)である。この基本規格から、応用別の機能と支援水準等を選択した機能規格(functional standards)は、応用仕様(application profile)と呼ばれる(図1)。



種々の応用仕様が、基本規格のうちの必須機能を備えていれば、異媒体間での情報交換の道が開かれる。

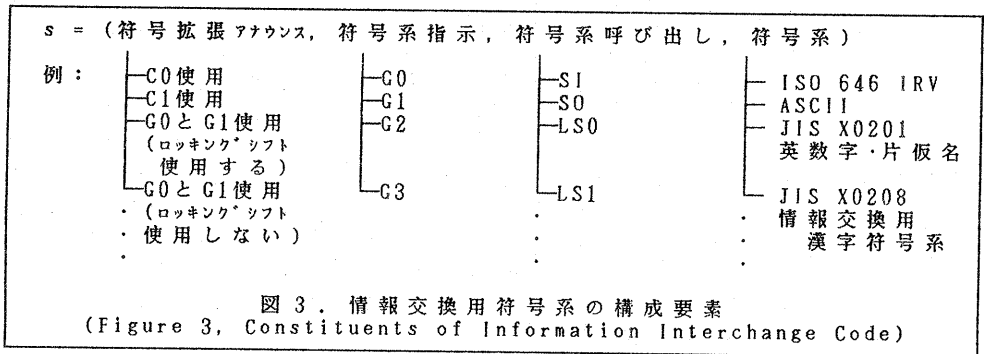
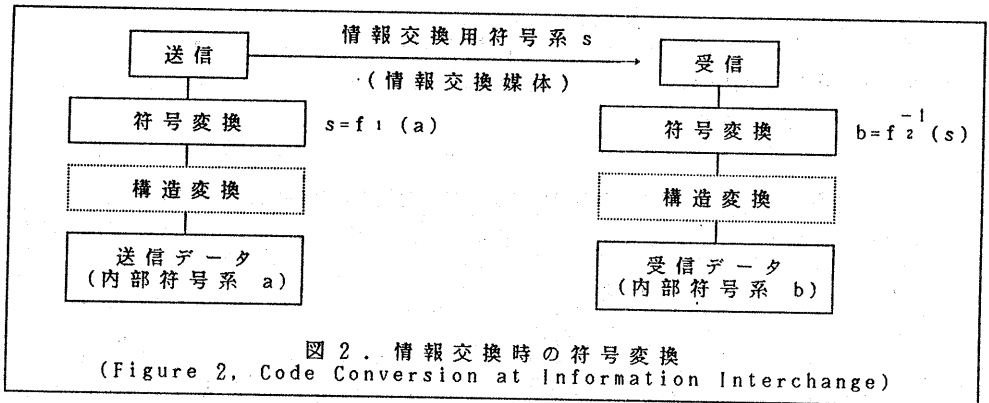
しかし、日本語の再現性を保持することを考えると、複数の応用仕様間での、日本語符号系の統一が必要である。

いま、図2のような情報交換を考える。情報交換媒体(通信回線・貯蔵媒体)の界面で、共通な情報交換用符号系 s を想定し、内部符号系 a の系統から送出する直前に、符号変換関数 f_1 により情報交換用符号系 s に変換し、受信側では内部符号系 b に変換関数 f_2 で逆変換する。

もし、同一内部符号系 ($a=b$) である特異な場合には、

$$f_1(s) = f_2(s) \quad \text{かつ} \quad f_1^{-1}(s) = f_2^{-1}(s)$$

である。
 ここで、情報交換用符号系 s は、符号拡張アナウンス、符号系指示シーケンス、符号系呼び出し並びに一つ又は複数の符号系によって構成される(図3)。



情報交換用符号系の構成要素の選択が同一でないと、変換関数及びその逆関数の組みを、複数組み備え、状況により使い分ける不便が生じる。

本稿は、広範囲の情報交換を容易にするために、将来の応用仕様で定める統一的な情報交換用符号系sの構成法を提案する。なお、本提案はさらに拡張機能アナウンスを加えて、符号拡張を制限するものではない。

2 現在の情報交換用漢文字符号系の扱い

現在の JIS X0208(83)情報交換用符号系は、その指示・呼び出しを図4の如く定めている[文献17]。

4. 符号

4.1 符号の単位 この規格で規定する符号の単位は、2バイトとする。各バイトは、JIS X0201(情報交換用符号)及びJIS X 0202による7単位又は8単位とする。先行の1バイトを第1バイト、後続の1バイトを第2バイトという。

4.2 2バイト符号の指示及び呼び出し

4.2.1 7単位符号系における指示及び呼び出し 3キャラクタのエスケープシーケンスである ESC 2/4 4/2で指示し、SIにより呼び出す。

4.2.2 8単位符号系における指示及び呼び出し 3キャラクタのエスケープシーケンスである ESC 02/4 04/2で指示し、呼び出す。

4.3 図形文字用符号 図形文字用符号は、表のとおりとする。各バイトのb7~b1は、7単位符号又は8単位符号のb7~b1に対応する。8単位符号では、b8は常にゼロとする。

図4. JIS X0208(83)の符号の規定(抜粋)
(Figure 4, Definition of JIS X0208(83))

すなわち、その指示と呼び出しは、7単位ではESC 2/4 4/2で指示し、SIで呼び出す。8単位ではESC 02/4 04/2で指示し、呼び出す。また、8単位図形文字用符号のb8は常にゼロとする規定がある。したがって、8単位で漢文字符号系と1バイト符号系を混用するには、図5の如く、符号拡張アナウンス・符号系指示・呼び出しを必要とする[文献16]。

a) 着信時の符号拡張アナウンス

ESC 02/0 04/7 C1 集合を使用

ESC 02/0 04/4 G0/G1を列02~07及び列10~15に呼び出す。
ロッキングソフトを使用しない

ESC 02/1 04/0 C0 に ISO 646を指示

ESC 02/2 04/3 C1 に ISO 6429を指示

b) 1バイト図形文字の指示

ESC 02/8 04/10 G0にJIS α -7文字を指示

ESC 02/9 04/9 G1にJIS片仮名を指示

c) 2バイト図形文字の指示

ESC 02/4 04/2 G0に日本語漢字を指示

例: 01/11 02/0 04/4 01/11 02/0 04/7 01/11 02/1 04/0 01/11 02/2 04/3
G0/G1使用 C1 使用 C0にISO 646指示 C1にISO 6429指示

01/11 02/9 04/10 03/4 04/1 01/11 02/4 04/2 03/4 04/1
JIS α -7文字指示 '4' 'A' 漢字指示 '漢'

図5. 現JISによるデータ列の例
(Figure 5, An example of data string in JIS X0208)

このデータ列は、エスケープシーケンスの分だけ、データ量が増加する。

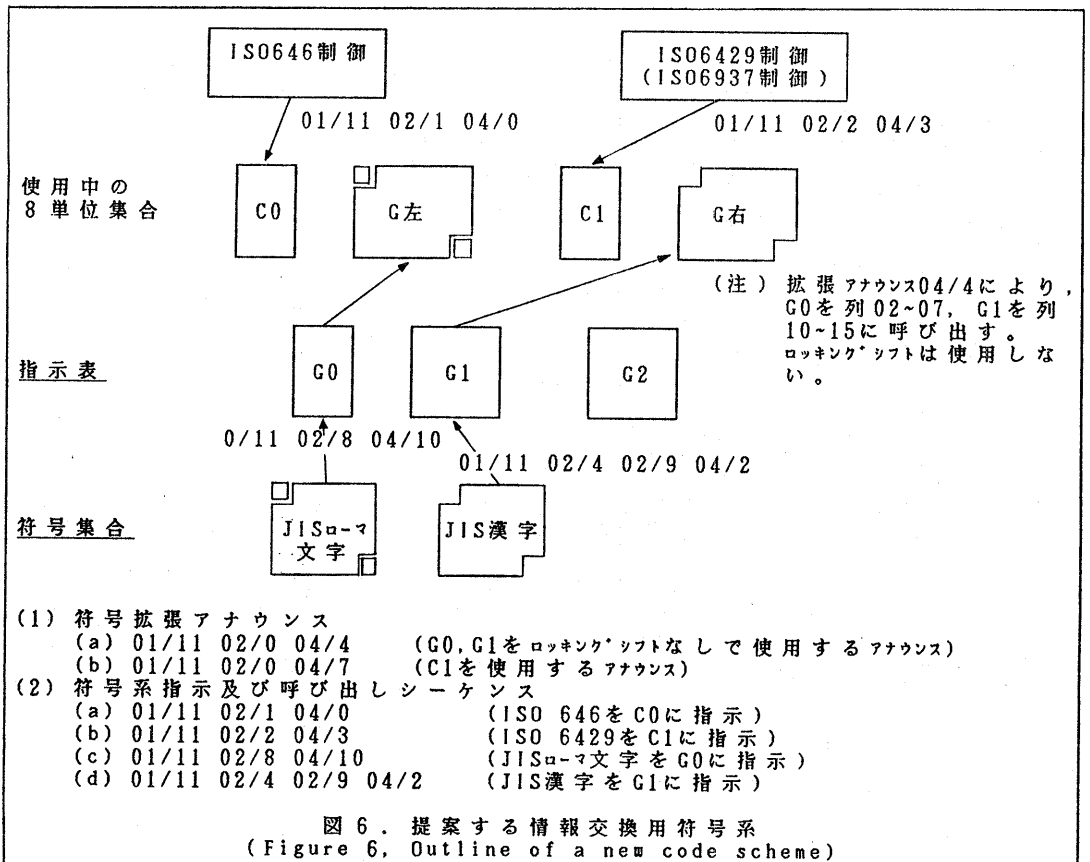
3 提案する情報交換用符号系の構成

ここで提案する構成方法は、文章中の符号系指示エスケープシーケンスを不要とし、データ量を減じるものである。ただし、図4のJIS X0208情報交換用漢字符号系の4章のうち、「エスケープシーケンスの終端文字を、G1の使用を認めるように拡張し」かつ「図形文字用符号のb0を'1'にしてもよい」と、1988年においてJIS X0208の見直し時点で拡張されることを前提とする。

提案する符号系の構成法の概念を図6に示す。

この方式で特徴的なのは、次の4点である。

- (1) 日本語漢字をG1に指示し、列10~15に呼び出す。
(ただし、JIS X0208の改定を前提とする)
- (2) JISローマ文字はG0に指示し、列02~07に呼び出す。
- (3) 符号拡張アナウンス 04/4により、G0を列02~07、G1を列10~15に呼び出し、ロッキングシフトを使用しないことを明示する。
- (4) 8単位及び7単位のデータ列は、C1に指示した機能符号の扱いと7単位でSI/SOを使用する点を除き、共通である。



尚、拡張アナウンス 04/4 の JIS X0202 の規定を図 7 に示す。

4/4 G0とG1集合を使用する。7単位系では、S1はG0を、S0はG1を、列2~7に呼び出す。8単位系では、G0及びG1集合を指示するエスケープシーケンスは、またG0及びG1集合をそれぞれ列02~07及び列10~15に呼び出す。ロッキングシフトは、使用しない。

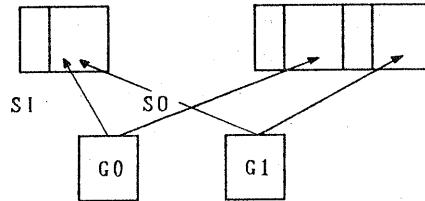


図 7. 拡張アナウンス
(Figure 7, Code extension announcer)

また、提案する符号系の構成法のデータ列見本を図 8 に示す。

イ) 符号拡張アナウンス及び指示

01/11 02/2 04/4 01/11 02/0 04/7 01/11 02/1 04/0
01/11 02/2 04/3 01/11 02/8 04/10 01/11 02/4 02/9 04/2

ロ) 書体制御なしの例

03/4 04/1 11/4 12/1 03/4 04/1.....
'4' 'A' '漢' '4' 'A'
符号系指示 エスケープシーケンスなし

ハ) 書体制御付きの例

S	S		G	S	
GP ₁	HP ₂	11/4 12/1	SP ₃	HP ₄	03/4 04/1
R	S	'漢'	R	S	'4' 'A'

ここで、P₁、P₃ は書体(全角,半角)を指示するパラメタ
P₂、P₄ は文字間隔指示のパラメタ

図 8. JIS X0208(88年改定後)のデータ列見本
(Figure 8, Example of proposed data-stream)

4 利点及び欠点

4.1 利点

ここで提案した情報交換用符号系の構成法の利点は、データ列が短い、漢字及び1バイトローマ文字データの判別が容易、国際間のローマ文字のデータ流通性が高いなどがある。また、この方法を操作系内部符号系として使えば、漢字領域でプログラム言語の構文誤りを誘発することを妨げる。例えば、マイコン用操作系の内部符号シフトジス(78)では、直輸入のProlog, Lisp, FORTRAN, Pascal言語処理系などで、字種により構文誤りがでていた。

4.2 欠点

一方、この欠点としては、将来のデータを変更する手間を必要とする点である。例えば、JIS片仮名文字はJIS漢字符号の片仮名に変更し、活字寸法縮小及び文字間隔縮小の制御符号を付加する。また、従来データの漢字符号は、第1バイト・第2バイトとも8ビット目を"1"にし、指示エスケープシーケンスを削除するなどの必要がある。しかし、事務文書体系の時代からは、従来暗黙的・便宜的に使用してきた「符号系と文字表現の寸法との関係」が破綻するから、文字ビットを観察しながらデータ変換を必要とするので、この符号拡張法だけの手間の増加とも言えない。

5 まとめ

次の世代の日本語文書交換の基盤符号系として、漢字符号をGI集合で使用する簡便な方法を提案した。この方式は、単に通信界面だけではなく、操作系内部符号系への応用も考えられる。短期的には、「JIS漢字をまた変更する」罪もあるが、データ列が短くなるので、この方式を実現できるようにJIS X0208を改定するのは、永い目で見ると良い方向であると信ずる。次世代へ向けて事務文書体系の実装規約を検討する組織(INTAPやTTC)での応用検討をお薦めする。

謝辞： 本件に関し、有益な討論をいただいた情報処理学会情報規格調査会日本語機能専門委員会委員各位、情報処理学会SC18専門委員会及びSC18/WG3.5合同委員会委員各位、日本規格協会テキスト交換システム調査研究委員会委員各位、及び日常種々の支援をいただく日本ユニバック(株)技術研究部各位、特に本論文の清書をしていただいた石川陸子氏に感謝する。

参考文献

- (1) 若鳥陸夫；"事務文書体系の紹介"，マイクロコンピュータ研究会，情報処理学会，情報研報MC42-1，(1986)
- (2) 日本規格協会；"システムソフトウェアの標準化に関する調査研究(テキスト交換システム)報告書"(1986)
- (3) 日本規格協会；"高度ネットワークのためのプロトコルの標準化に関する調査研究(テキスト交換システム)昭和61年度報告書"，(1987)
- (4) 若鳥陸夫；"事務文書体系(ODA)の規格化とハフマンコーディングシステム"，画像電子学会予稿，86-06-03，(1987)
- (5) 春田勝彦；"文書交換と転送"，情報処理，Vol. 28, No. 4, pp505-pp509, 情報処理学会，(1987)
- (6) ISO；"DIS8613/1.2, General introduction, ODA"，(1987)
- (7) ISO；"DIS8613/2.2, Document structure, ODA"，(1987)
- (8) ISO；"DIS8613/4.2, Document profile, ODA"，(1987)
- (9) ISO；"DIS8613/5.2, Office document interchange format(ODIF)"，(1987)
- (10) ISO；"DIS8613/6.2, Character content architecture, ODA"，(1987)
- (11) ISO；"DP8613/7, Raster graphic content architecture, ODA"，(1987)
- (12) ISO；"DP8613/8, Geometric graphic content architecture, ODA"，(1987)
- (13) 若鳥陸夫；"日本字フォントフォーマット向け「事務文書体系」の「応用仕様」" 情報研報MC45-2, マイクロコンピュータ研究会，情報処理学会，(1987)
- (14) 若鳥陸夫；"事務文書体系の国際規格案の紹介"，OS87-7, 電子情報通信学会，1987
- (15) 日本工業標準調査会；"JIS X0201, 情報交換用符号"，日本規格協会，1976
- (16) 日本工業標準調査会；"JIS X0202, 情報交換用符号の拡張法"，日本規格協会，1984
- (17) 日本工業標準調査会；"JIS X0208, 情報交換用漢字符号系"，日本規格協会，1983

- 以上 -