情報システム 16－9
（1987. 9. 22）

国際規格と文化圏の問題

－国際規格に漢字処理を付加する苦労－

若鳥陸夫

（日本ユニバック(株)）

筆者等は，ISO/TC97/SC18「文章及び事務系」で，国際規格制定作業中の「事務文書体系(ODA)」の国際規格に，漢字処理を組み入れる努力をここ6年間行ってきた。しかし，「日本国の漢字はその全集合を確定できない」，「類似漢字字形の区別があいまいである」，「日本国内の事務所では，手書き図形文字時代から活字文明時代への転換期にある」などの問題がある。それ故に日本文字では，自然発生した図形文字をある中央値にまとめる努力の必要性が生じている。本報告では，その前時代的性質を持つ漢字を，国際規格に取り入れる際に，国内外で苦労した問題について報告する。

Some Difficulties between International Standardization and National Culture

- Including of Japanese processing to ISO DIS8613 -

Rikuo WAKATORI

Nippon Univac Kaisha Ltd,
2-17-51, Akasaka, Minatoku, Tokyo, 107, JAPAN

Author reports on some defficulties to incorporate Chinese kanji characters into International Standard has been discussed in ISO for six years.

1．序論

漢字は，象形文字に端を発して，その意味上の素片を，２次元的に組み合せて，別意味の漢字を生成するという具合に，表意文字として発展してきた。新素材・新現象・新家庭などの毎に，新しい漢字が創造され，図形の識別で用をなしてきた。その内，読み難い漢字や類似文字が増加してくると，整理統合が行われ，表音文字での代替も行われた。しかし，日本国では，人名の識別としての「姓」について，「図形」の一致という政策(戸籍法)をしてきたため，誤字・俗字や難読文字も「姓」として正式に流通している。近隣諸国の漢字文化圏との交流に際して，新漢字が輸入されるので，一国の国語政策だけではその整理は片付かない。さらに，活字(タイプライタ)での扱いは，多字種の入力手段での実現に労力が向けられ，欧米より数十年遅れてようやくワードプロセッサが普及し始めた。

印字出力も，小面積で微細な区別をさせる必要性と，字種の多さで複数書体を具備するには，多大の資源を必要とするところから，現在の情報処理系では，日本で「明朝体」と呼ばれる書体群とパイカ体くらいの取扱いが一般的である。

一方，局所的な応用では，レーザプリンタと組み合せて，高品位・多書体の印字出力が可能となり，卓上出版といった応用も要求が強い。しかし，「開放型システム間相互接続(ＯＳＩ)」を基本にした，２１世紀の広域での不特定間での情報交換を考えると，その字種は現在の日本工業規格「情報交換用漢字符号系」に止まらず。もう少し広い範囲の字種の要望も強いし，書体・寸法・字間などの可変性の実現の必要性もある。これらは，単に，情報処理系で取り扱う漢字種類の拡大ですむ問題ではなく，近隣諸国を含めて，100年レンジで法務省水準での姓名用漢字の統合化も必要であろう。

2．国際規格の概要

我々(ISO/TC97/SC18)が国際規格制定作業中の事務文書体系(Office Document Architecture，略称ＯＤＡ）は，次世代の文書構造や文書処理の規則集である。その主な特徴は次の如くである[文献1..11]。

（１）国際及び国内の文書交換
（２）自動編集や自動割付け処理
（３）再処理可能な「執筆意図（論理構造）」や「割付け結果（割付け構造）」を含むデータ構造
（４）共通文書群に対して，テンプレート（共通構造）を生成し，そのテンプレートから特定文書を生成できる
（５）論理構造，割付け構造，表現体裁，割付け体裁などからなる構造
（６）文字，ラスタ図形，幾何学的図形が混合使用できる
（７）豊富な書体制御（無限／１文書，１０種／基本対象体）

（8）文字寸法指定，文字間隔指定は25.4ミリ/1200を基本測定単位とする
（9）表現指定
下線，二重下線，反転，非反転，イタリック体，高輝度，
低輝度，太字，細字など
次にこれらの特徴を個別に解説する。

2.1 国際・国内の文書交換

事務文書体系と従来形式の文書交換との極端な違いは，「開放系」か「閉路系」である。前者は，国境，製造者の境界，操作系の境界を越えて，「どこでも」・「だれとでも」文書交換する体系なのに対し，後者は，同一製造者の同一操作系の文書を主な交換対象としている。両者間には，利用目的の違いもあり，一概に前者ばかりが良いわけではなく，文書交換の界面では「事務文書体系」であるが，特定な系統内では，後者の様々な水準のものが接続され，その系の規模と情報交換の必要性，さらには局所的印刷制御の必要性などにより，利用形態が分かれよう。
しかし，5年から10年レンジの機器更新時期ごとに，新旧混合の利用形態でも次第に前者の比率を高めていくことになるとみる。
この事務文書の交換のために，「事務文書交換様式(ODIF)」が規定されている。

2.2 自動割付け処理

編集処理の（内容編集及び構造編集）の終了した文書は，共通割付け構造をテンプレートとして，割り付け体裁指定を合わせ，自動的に割り付けることができる。

2.3 再処理可能なデータ構造

文書の章立てなどの論理構造は，ページに割付けされて「割付け構造」を生成する。一般的な割付け処理では，論理構造に割付けのための記述を追加し，論理的な執筆者の意図を書式で代替してしまう。それを再編集処理する際には，割り付けられた状態で主に手作業で再編集を行うことになる例が多い。

ところが，事務文書体系では，再処理を自動的に行うために，割付け処理後にも論理構造を残し，付加した書式符号を一括して無効にできるデータ形式がある(図1)。
日本語の振り仮名の例を図2に示す。

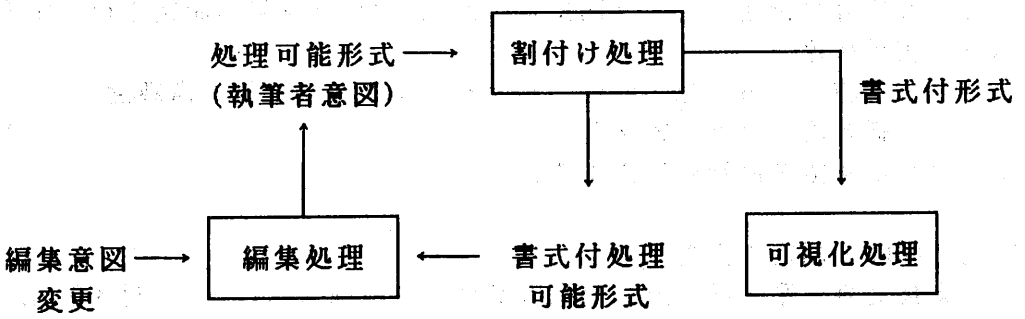

図1　事務文書体系の書式付処理可能形式の再編集

2. 4　共通文書構造から特定文書構造の生成

テンプレートとなる共通文書構造は，論理構造と割付け構造の双方にあり，それぞれ，「共通論理構造」及び「共通割付け構造」と呼ばれる。この共通文書構造から特定の文書構造を生成できる。
これらの構造は，個別でも全体でも交換することができる。
もし，共通文書構造を交換すれば，「転送効率改善」，「文書間の整合性向上」，「原作者の意図にもとづき対象体や文書を生成できる」などの利益をもたらす。

2. 5　内容体系の種類

「基本対象体」及び「区画」は，「文字」・「ラスタ図形」・「幾何学的図形」のうち，一つの内容体系になる。
すなわち，「ページ」又は「枠」は，文字・図形の組合わせで構成できる。「文字内容体系」は，文字(英数字，漢字，その他記号)と機能符号との列で構成される。「ラスタ図形内容体系」は，グループ3及び4のファクシミリの体系で，その内の一つの書式形式はグループ3及び4のファクシミリに出力できる形式である。
「幾何学的図形内容体系」は，コンピュータ図形をGraphic Metafile形式(ISO/TC97/SC21担当) で表現する形式である。

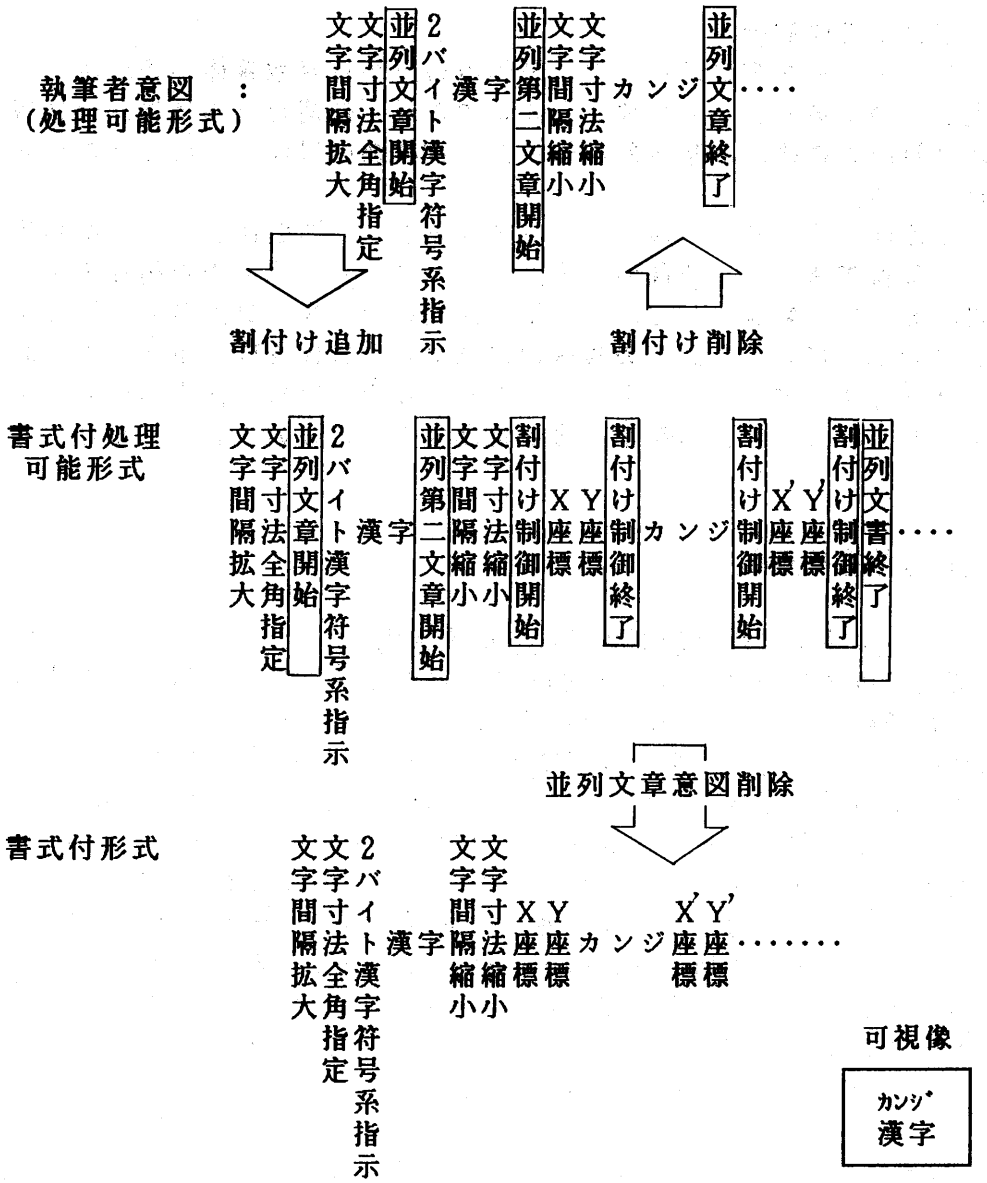

図２　処理可能形式／書式付処理可能形式／書式付形式

2. 6　豊富な書体制御

コンピュータ文書処理は，単一書体の応用から複数書体への応用へと拡大されつつある。この事務文書体系での最上位適合水準では，一文書あたり無限の書体，１基本対象体あたり10種類の書体（寸法を含む）を指定できる。

たとえば，「エリート体」，「パイカ体」，「明朝体」，「丸明朝体」，「教科書体」，「新聞活字体」，「丸ゴシック体」の書体仕様を「文書概要」に一覧表としておき，基本対象体で「書体識別番号と縦寸法」との対応を指示し，図形表現選択機能により書体を呼び出す。文字間隔・文字寸法・書体・符号系がそれぞれ別要素のため，従来の暗黙使用であった「符号系による活字寸法の自動変更」の扱いに終止符が打たれる。

3．国際規格と漢字使用圏

3.1　あいまいな漢字集合と識別番号付与の困難性

情報交換用漢字符号系(JIS X0208)の範囲の字種だけを使用した文書の交換では問題とならないが,その符号表にない姓名や地名の伝達には，局所的に生成した「外字」もその対象としたい。しかもその外字も含めて，一連の識別番号を付与したい。しかし，それらの字形の一部は,図形としては識別できるが，ある字と別な字との間は，実数間隔で変形した字形が存在する可能性を秘めているため，ある人の目前にある漢字集合によって，一連の番号を付与できない性質のものである。その識別が広域で一意でないと，広域の情報交換の用を足さないので，あえて識別子を付与するとすると，実数表現で 1.0125というように，ある基本形にどれだけ近似しているか測定基準を設けて，あらわす方法を思い付く。
しかし，この方法は，大きな支障がある。

①漢字を区分し，実数表現にする作業をどうするか？
②漢字種類の拡大の際に整合性を保証できるか？
③世界的に，実数表現(ｽｹｰﾘﾝｸﾞを含む)の符号系が受容できるか
（桁数,離散配置）

もっとも悪いのは，どんな漢字の出現も体系上許容し，混乱の増大を積極的に支援することにあるだろう。
国際標準化機構(ISO)では，現在書体の登録とその識別番号の付与を検討中である。しかし，日本国を始め，中国・韓国などの漢字圏では，漢字の全集合を特定できないし，将来の拡張を前提にした識別番号を公式に割り付けることはできそうにない。

3.2 外字の存在とその統一識別番号付与の困難性

広域の情報交換に，情報交換用漢字符号系以外の外字の使用是非は論じないこととして，もしその外字を情報交換に使用するとしたら，図形として扱うか，存在する外字を収集して識別番号を付与するかである。後者は，情報交換用漢字符号系以外の外字(誤字も含む)を国家ですら完全収集できてはいない。これができないと，流通しているすべての漢字に識別番号を付与したことにならないから，又何らか便宜的手段を必要とする。

3.3 多字種と登録の困難性

国際的に各字の書体を登録したいという話がある。字種の少ない言語圏では，複数書体の登録によっても，保管される原版の枚数は管理できる水準であろうが，日本のように，各出版社等が各自の書体を1万字×15種類位は保有していると想定される状態では，その一元管理はとうてい不可能である。我々日本語を自由に使用できる人間ですら，一万字の「照合」と「並べ替え」は大変な労力である。その図形文字の自動照合のアルゴリズムを発見しない限り，機械処理もできない。

我々が可能なのは，書体群(明朝体,教科書体など)への識別番号を与えるのが第1歩である。ISO/TC97/SC18では，一字ごとの書体を登録管理することを検討しているが，日本としてはその方法では当面保証し兼ねると発言している。しかし，自国の文字集合の境界を識別できないことを理解させるに至っていない。それほどまでに，日本国は後進国(エジプト時代相当)とは信じられないのも無理はない。

3.4 手書き文化の押しつけ→漢字への振り仮名

多字種言語で，かつ一字に複数の読み方がある言語にとって，振り仮名の付与が必要である。その振り仮名は「漢字(ｶﾝｼﾞ)」のように後続してもよいし「漢字(カンジ)」のように，2行にして書く場合とがある。

前者は，「行を主体とするタイプライタの発達した国国での流儀」，後者は「タイプライタ後進国での手書き文化の遺物」という冷静な見方をすると，ワードプロセッサの普及の後では，日本語も100年レンジで前者の流儀へ進化していくと推定される。

その先進的か後進的かの主観論はさておいて，後者の流儀での振り仮名(ﾙﾋﾞ)の付与を国際提案して1年がかりで討論し了承された。

その審議途中で，問題となったのは，その必要性もさることながら，事務文書体系でいう「処理可能形式(割付けられていない論理構造)」，「書式付形式(紙に割り付けられた状態)」,「書式付処理可能形式」とで,それぞれ振り仮名を「編集処理」，「割付け処理」及び「可視化処理」でどう取り扱かうというものであった。

3.5 半角・全角の問題

文字の寸法も，日本での応用が混乱しているものの一つである。それは，欧米からの英字・数字のパイカの寸法と，2バイト系漢字符号の図形表現を約2倍にした局地的便宜的実現にある。すなわち，「符号系では，文字の大きさは定まらない」という原則によらず，暗黙的に活字寸法を符号系の属性としてしまった実現が多い。

従って，テレテックスまでの応用では，ある寸法の全角から半角へ，書体と活字寸法及び文字間隔を一つの機能符号(GSM)で処理してしまっていた。ところが，それは最近の，「符号系」・「寸法」・「文字間隔」・「書体」の独立制御，並びに「処理可能形式」，「書式付形式」，「書式付処理可能形式」の三形式を統合して考える新しい流儀とは，合致しなくなってきた。

また，情報処理系の出力する印字の寸法は，目下のところ，印刷のポイントとの一意の対応性はない。

ところが寸法を自由に設定できる装置が出現し，事務文書体系で「完全一致再現」を求める時代になると，各自の勝手な基準寸法を「全角」と呼称している現在は，また，一時的解決であることが見えてくる。

3.6 日本の国際会議での役目

国際会議では，欧米での1バイト符号系での取り扱い文字の範囲が当面の関心であり，2バイト符号系での取り扱い可否の眼は日本からの委員にしかない。例えば，図形文字を使用可能の記述は，英数字だけなのか漢字も含むかの質問をすると，2バイト符号系を忘れていた事態に気付くといった状態である。国際規格だから，全世界の言語や文字を取り扱い対象とした総論は多いが，各論で2バイト符号系での振る舞いについて正確に記述した作業案(WD)は少ない。また，日本からの注文で，基本思想の転換を求めることもあるので，平行線のまま，規格草案(DP)になるものもある。日本国の委員は，多字種国の代表者であるとの認識に立ち，多字種を取り扱える国際規格の制定に初期段階から寄与する必要がある。

3.7 国内での問題

3.7.1 内部符号系対情報交換用漢字符号系

世界水準での文書交換を考える場合，不特定者間では，世界で一つの符号体系で送受信しなければならないのが当然である。筆者は操作系内部での，内部漢字符号系を拒否するものではないが，通信路の符号系を「情報交換用漢字符号系」に一本化すれば，相互接続可能な第1水準になると考えている。 従って，ワードプロセッサ・

マイクロコンピュータから，大型コンピュータに至るまで，通信回線出入口に「情報交換用漢字符号系」の変換機能があれば，通信路の両端は，たとえ内部符号系であっても，異操作系や異機種との相互接続性は格段に向上する。また，全ての金物において，この考えに沿う準備をすることが次の世代へ残す我々の遺産であろう。
通信路上だけが，情報交換用漢字符号系であるのなら，編集系での漢字・非漢字の判定などの方法は，３バイトエスケープシーケンスを探索するなど不要であり，その弁別はその操作系にまかされる。

3.7.2 機能符号と図形文字

２バイト図形文字区間で，１バイト機能符号，例えば復帰(CR),改行(LF)，間隔(SP)の受信のできない軟件の実現がある。その受信処理は，２バイトの漢字符号系扱いしてしまい，誤符号と検出してしまうのである。それは機能符号群[C0集合]と漢字符号(ﾏﾙﾁﾊﾞｲﾄC0集合)との並列受信の原則を忘れた実現である。
パーソナルコンピュータにこの実例が多く，一行毎に漢字脱出をしなければＣＲやＬＦを受信できない大手コンピュータメーカの軟件もある。
間隔(SP)の日本国内での実現は，偏差が大きい。間隔は「図形文字とも機能符号とも両方に見なせる」というＪＩＳの記述の安全側解釈をしていないことによる。どちらにでもなるなら，安全側の機能符号として，漢字モード中でも受信処理するようにするのが正解である。さらに，その実際の間隔表現はその区間の文字幅指定の通りとするのである。

まとめ

事務文書体系(ODA)の国際規格の制定作業並びにパーソナルコンピュータによる電子掲示板を運用していて気付いた情報交換の困難性の一面を報告した。事務文書だけでなく，プログラム言語の国際規格での統一的な日本語の扱いが必要だし，今後はより高位水準の文書交換が国際間で必要になる。国際間で電子的に編集・校正・製版を分業できるようにするためには，多字種国がもっと国際規格の場で活発に活動しなければならない。今や国際規格制定後各国語を付加してすむ水準の国内規格を制定する時代ではなくなっている。また，高位の漢字表現・文書交換には，今後，２１世紀への遺産を残すつもりで広範囲の分野の方々の活動が不可欠である。

謝辞：本件に関し，日頃有意義な討論をいただく，情報処理学会情報規格調査会技術委員会SC18専門委員会委員各位，同SC18/WG3･5合同委員会委

員各位，ISO/TC97/SC18/WG3 英米仏独伊蘭加国の委員各位，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，筆者の後方から支援をいただく日本ユニバック(株)技術研究部員各位，並びに本論文の清書をしていただいた石川睦子氏に感謝の意を表わす。

参考文献

(1) ISO;"DIS8613/1, - Information processing - Text and office system - Office Document Architecture(ODA) and interchange format - Part 1: General introduction",(1986-09-18)
(2) ISO; "DIS8613/2,- Information processing - Text and office system - Office Document Architecture(ODA) and interchange format -Part 2: Document structures",(1986-09-18)
(3) ISO; "DIS8613/3, -Information processing - Text and office system - Office Document Architecture(ODA) and interchange format - Part3: Document processing reference model",(1986-09-18)
(4) ISO;"DIS8613/4, - Information processing - Text and office system - Office Document Architecture(ODA) and interchange format - Part4: Document profile",(1986-09-18)
(5) ISO;"DIS8613/5, - Information processing - Text and office system - Office Document Architecture(ODA) and interchange format - Part5: Office document interchange format(ODIF)",(1986-09-18)
(6) ISO;"DIS8613/6, - Information processing - Text and office system -Office Document Architecture(ODA) and interchange format - Part6: Character content architecture",(1986-09-18)
(7) 若鳥；"事務文書体系の紹介"，情報処理学会，マイクロコンピュータ研究会，情報研報MC42-1,(1986)
(8) 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)
(9) 日本規格協会;"高度ﾈｯﾄﾜｰｸのためのﾌﾟ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)報告書,(1987)
(10) 若鳥;"事務文書体系(ODA)の規格とﾃﾞｽｸﾄｯﾌﾟﾊﾟﾌﾞﾘｯｼﾝｸﾞ,画像電子学会予稿 86-06-03,(1987)
(11) 若鳥;"事務文書体系の紹介(ﾁｭｰﾄﾘｱﾙ),ｵﾌｨｽｼｽﾃﾑ研究会予稿,電子情報通信学会,(1987)