マイクロコンピュータと
ワークステーション 55－5
（1989. 5. 16）

言語変換系模型cps
― ワードプロセッサからページ記述言語への変換方式の提案 ―

若鳥陸夫
日本ユニシス株式会社
107，東京都港区赤坂2－17－51

梗概

文書作成編集系の文字列出力から，ページ記述言語（例えばPost Script［文献1］）への変換の中間言語の方式の提案を行なう。中間言語をうまく定義すれば，ページ記述言語を変更してもその中間言語から特定言語への変換系を準備すればよいのでファイル形式を共通にできる。この提案では，中間形式の文字列をGR漢字符号とし，文書を表示装置へ出力するとき，その表示装置内部符号変換プログラムをデータ列に前置する方式とする。このような前置方式にすれば，計算機内のデータ形式は，その表示装置に依存しない形式の文字列となるので，汎用性がある。

A Proposal of Intermediate Language between Japanese Word-Processors and Page Description Languages

Rikuo Wakatori
Nihon Unisys Limited
2-17-51, Akasaka, Minato-ku, Tokyo, 107, JAPAN
E-mail: wakatori@s7040.unisys.junet

The author proposes an intermediate language (cps) bridges between word-processors and page description language (PDL). This keeps compatibility in the output data stream, especially in multi byte Kanji character codes at file system.
A key feature of this proposal is that a document prolog consists of short form of PostScript language and code conversion rules from two byte interchange code to internal code in a representation device. The document prolog is sent to a presentation device with document body.
NOTE: All descriptions were written in Japanese, but English version will be supported by the author if requested.

1．研究の位置付け（purpose of study）

ページ記述言語は，ページ記述言語で表現した情報を，印刷装置独立な文書表現形式として，ページ記述言語の表現力を，印刷装置の出力装置の特性に関係なく，一定にする。

ページ記述言語は，ページ記述言語で表現した情報

f）文字列表示指示　　文字列　＿s
g）文字寸法指示　　ポイント　＿sc
h）ページ表示指示　　＿sp
i）座標変換指示　　座標　＿t
j）直線描写　　座標群　＿l

2）符号系変換部（code conversion rules）
符号変換部は，文書に前置して印刷装置側で動作する符号変換プログラムからなる。この符号系変換部には，次の3機能が入っている。
a）7ビット対8ビット変換（octet conversion）
通信路が7ビット系のとき，SO（0／14）で列10から列15に呼び出し，SI（0／15）で列2から列7に呼び出す。これらの制御符号が出現した後は，再びこれらの制御符号で状態を変更されるか，文書印刷になるまで持続する。また初期値は，列2から列7に呼び出されているとする。
b）情報交換用漢字符号系の弁別（GR detector）
英字符号（JIS X 0201ローマ文字または米国情報交換符号系ASCII）と情報交換用漢字符号系を弁別する。ここでは，漢字符号が，列10から列15に呼び出される場合を試験した。この符号呼び出し方法では，最上位ビットの状態で漢字符号か英字符号を識別できる。
c）漢字符号系の内部表現への変換（conversion）
情報交換用漢字符号系による2バイト漢字符号を，装置固有の符号（例えば4バイトのグリフ）に変換する。この変換により，装置固有の字形識別子をその装置内に封じ込めることができる。

3．実験方法（test method）

前置部を自動的に生成する処理系の模型を，Pascal言語で書き個人計算機ユニシスUP10E50で翻訳し実験した。言語処理系は，Turbo－Pascalで，原始プログラム量は750行，36kbであった。その後，ユニシスNUL J－3100GTや東芝J－3100SGTへ移植し，使用環境を拡大した。

4．評価（results）

4．1．ファイルの共通性（file compatibility）
この方式によれば，前置部は装置依存性を記述しているが，文書本体部の様式は出力装置に非依存の形式が実現できる。これは，あたりまえのことであるが，装置の造りが多様化したときにこの思想が真価を発揮するであろう。
この方式では，印刷装置出力の前の漢字符号系が，情報交換用漢字符号系に統一され，一般の表示装置で容易に可視化できる。

4．2．適応の拡大性（applicability）
この前置部により出力装置依存部を吸収する方式は，PostScript言語だけでなく，他のページ記述言語へも適用できる。前置した言語変換部を対象のページ記述言語に変更すれば，対応できるからである。ここでは，漢字符号系としてGR漢字方式を採用したが，変換プログラムさえ用意すれば個人計算機の操作系で多用されているシフトジス（78）及びシフトジス（83）にも適用可能である。

4．3．グリフ出力装置の救済対策（advice for Glyph）

この研究は，印刷装置と計算機の界面に，情報交換用漢字符号系でなく，独自のグリフを使用した装置の救済に見える。しかし，これは単にPostScript代わりに搭載した印刷装置の使い方を示したのではなく，文字符号の代わりに字形識別子（グリフ）を使用したときのファイルの共通性の確保を試験していることに相当する。

国際標準化機構（ISO）で標準化審議中のグリフ（ISO DIS 9541）による界面を可視状態に維持したい応用のとき，この方式の適用が必要な場面が増加するかも知れない。

5．あとがき（summary）

ページ記述言語は，日本語文書作成系や各種の応用界面の交換様式として利用される機運にあるが，編集系形式の報告で述べたように，印刷装置内部形式を統一文書の中で出力すれば，計算機出力ファイル様式の変換則を前置情報とし，ページ記述言語のオリジナル目的に応用広く処理途中でとりかえる利便なら，ページ記述言語研究の目的は，広範な情報交換の思われる。

この報告書では，事項を何人でも使用することの利便の追求にあり，筆者は，ページ記述言語の交換方式の自由な使用を許容し，ページ記述言語の報告書に述べた文書の高性能を願っている。

この報告書では，単に PostScript 組み込み印刷装置の使い方ではなく，その界面の変形した文字符号を逆用して，将来出現するであろうグリフ（字形情報識別子）への対応法を試行したのである。

今後の研究は，書体と符号系の概念の分離していない印刷機構に，同じ書体を呼び出して，1バイト符号でも2バイト符号でも，同じ寸法の字形が表現されるような論理を追加することやグリフ表現の字形を編集系（例えば Emacs）で表示できるようにすることがある。前者は，漢字字形の呼び出しキーとして第一バイトの値を利用するのでは，文字概念の等価な文字に対して符号系が相違し，同じ字形を参照できるように基本的な事項である。これは，符号系は特定の字形や文字大きさと無関係の国際規格の基本的な符号系の方向や日本語書体情報交換，Emacs などに反映できる方法案については，概念に戻すことである。後者は結果を見て，残存する課題に対する解決方法案については，研究が進展したところ機会があれば報告したい。

参考文献（references）

（1） Adobe System inc.: PostScript Language Reference Manual, Addison-Wesley, Reading, Massachusetts, 1985

（2） ISO 8879: Information Processing — Text Composition — Standard Generalized Markup Language (SGML), 1986

（3） 若鳥陸夫 Prologによる標準ページ記述言語SPDLの早期試作，第54回マイクロコンピュータとワークステーション研究会予稿，情報処理学会，1989

（4） 日本規格協会：JIS素案「標準汎用マーク付け言語SGML」，マルチメディア調査研究報告書，1989年

付録（appendix）

言語変換処理系の概要（outline of prototype）

ここで作成した模型プログラムは，この報告の主体ではないが，参考のためにその概要仕様（抜粋）を示す。

a1．言語変換系の特徴（features of prototype）

この言語変換系の特徴は，個人計算機や文書作成編集系の出力として得られる文字列文書に，利用者がこの仕様書で規定した付せんを追加すると，言語変換処理系がPostScript言語様式に自動的に変換してくれることにある。
従って，利用者は言語PostScriptの知識を持たなくてもPostScript組み込みの表示装置を利用して高品位の文書を生成できる手段を提供している。ここで，付せん指定できる機能の概要は，次表の通りである。なお表記法は，標準汎用マーク付け言語［文献4］に近似させた。

項目	機能名称	付せん開始字句	備考
（1）	書体指定	＜／f	15種（装置依存）
（2）	文字寸法指定	＜／s	ポイント指定
（3）	文字修飾指定		
	下線	＜／ul	
	上付き文字	＜／u	書体，寸法指定併用可
	下付き文字	＜／d	書体，寸法指定併用可
（4）	表示媒体方向の指定		省略時は肖像
	肖像	＜／o portrait	縦長方向
	風景	＜／o landscape	横長方向
（5）	ページ表示指定		
	ページ表示開始	＜／p number	数字　自動加算
	改ページ	＜／p new	自動加算

この処理系の利用者界面は，編集系で直接付せん開始字句列を記述してもよいが，日常よく使用する付せん字句を機能キーなどに割付けておき，利用されることを意図している。なお，付せん終了の字句は”／＞”であり，付せん開始から付せん終了までを付せん区画と呼ぶ。ある付せん区画内で，再び付せんを使用してもよい。

a2．仕様（specification of prototype）

a2．1．入力データ様式（input data format）

この言語変換系の入力データ様式は，文字，文字列，復帰及び改行から構成する可視可能なデータ様式とし，原始文書及び付せんから構成する。いずれの場合も1行の末尾は，復帰及び改行で終端しているものとする。データ中の文字集合の範囲は，JIS X 0201情報交換用符号の列2－列7，JIS X 0208情報交換用漢字符号系の非漢字及び第一水準漢字とする。ただし，装置が第二水準漢字を実装しているときには，その使用を制限しない。

a２．２．出力データ様式（output data format）

言語変換系の出力データ様式は，中間言語cpsの構文様式とし，PostScript言語処理系内蔵の印刷装置に出力すると印刷できる様式とする。

a２．３．言語仕様（language specification）

a２．３．１．入力構文（syntax）

この言語変換系の入力構文は，次のように規定する。

```
(1)  入力構文        ::= { <原始文並び> | <付せん区画> }*
(2)  <原始文並び>    ::= [ <原始文> ]*
(3)  <原始文>        ::= 空文 | 行末がCR／LFの文
(4)  <付せん区画>    ::= <付せん区画開始>
                         <原始文並び> <付せん区画終了>
(5)  <付せん区画開始>::= <! <付せん>
(6)  <付せん>        ::= { <書体付せん>        |
                           <文字寸法付せん>    |
                           <文字修飾付せん>    |
                           <ページ制御付せん>  |
                           <媒体方向> }
(7)  <付せん区画終了>::= />
(8)  <書体付せん>    ::= f <書体名>
(9)  <書体名>        ::=
                 Times-Roman |
                 Times-Bold |
                 Times-Italic |
                 Times-BoldItalic |
                 Helvetica |
                 Helvetica-Bold |
                 Helvetica-Oblique |
                 Helvetica-BoldOblique |
                 Courier |
                 Courier-Bold |
                 Courier-Oblique |
                 Courier-BoldOblique |
                 Symbol
                 Ryumin-Light-Ext-H
                 Ryumin-Gossic-BBB-Medium
(10) <文字寸法付せん>::= s <ポイント>
(11) <ポイント>      ::= 1 | ... | 40
(12) <文字修飾付せん>::= <文字修飾種別>
(13) <文字修飾種別>  ::= <下線付き文字> |
                         <上付き文字> |
                         <下付き文字>
(14) <下線付き文字>  ::= ul
(15) <上付き文字>    ::= u <移動量>
(16) <下つき文字>    ::= d <移動量>
(17) <移動量>        ::= 任意の正負の整数
(18) <ページ制御付せん>::= p <ページ制御種別>
```

```
(19) <ページ制御種別>   ::= <ページ替え> | <ページ番号付与>
(20) <ページ替え>       ::= new
(21) <ページ番号付与>   ::= number <開始値>
(22) <開始値>           ::= 任意の正整数
(23) <媒体方向>         ::= o <方向>
(24) <方向>             ::= <肖像> | <風景>
(25) <肖像>             ::= portrait
(26) <風景>             ::= landscape
```

備考 1．この変換言語は，付せん区画での再帰的呼び出しを許す。再帰的付せん区画の終了指示は，一番内側の付せん区画を終了させる。
2．書体名とその種別は，印刷装置の実現に依存する。

a3．コード生成規則（code generation rules）

処理系は，入力文書を前方から順次走査して，字句の出現する状態によりコード生成を行なう。なお，目的言語は後置記法で記述するので，操作子を一番右側に，その引き数群をその左側に，それぞれ記述する。ここで使用した約10種類の操作子に関する限りでは，引き数は0個から2個である。

a3．1 文書概要のコード生成規則（rules for prolog）

コード生成が開始したとき，文書概要（prolog）を注釈文で記述し，定義群を記述し文書中では目的言語の操作子を短縮名（中間言語）で記述できるように宣言する。次に座標を左下（0，0），表現開始位置を左上（左余白，上余白）とし，書体名，文字寸法及び媒体方向の省略時想定値を指定する。これらのため，次のコード（抜粋）を生成する。

```
%!  convertPS          <版数>  <著作権>      PS慣習
%%Title:               <文書名>              PS慣習
%%Creator:             <文書作成者>          PS慣習
%%CreationDate:        <文書作成日付け>      PS慣習
   %%  Default  Font              =/Helvetica
   %%  Default  Font  Size        =12
   %%  Default  LineHeight        =14
   %%  Default  Orientation       =portrait
   %%  User  Unit                 =point
   %%                Page  Width  =595points
   %%                Page  Length =841points
%%EndComments                     PS慣習
%   Begin  Definitions            定義群開始
   /_f    findfont  def           書体探索
   /_m    moveto    def           絶対位置移動
   /_l    lineto    def           直線描画
   /_r    rotate    def           回転
   /_rm   rmoveto   def           相対移動
   /_s    show      def           文字列表示
   /_sc   scalefont  setfont  def 文字寸法指定
   /_sp   showpage  def           ページ表示
   /_t    translate  def          座標変換
%  End  of  Definitions           定義群終了
   Helvetica  _f                  省略時書体
```

```
        12   _s                                  省略時寸法
      0   0   _t                                 原点（0，0）
    <LM>  <UM>   _m                               左上隅へ移動
%%EndProlog                                       文書概要終了
```

ここに，注釈の頭文字列”％！”は，目的言語PostScriptの適合プログラムを示し，文字列”％％”は論理的構造の予約語を示す慣習である。<版>は言語変換系cpsの発行の名称（版），<著作権>は”Copyright 1989 Rikuo Wakatori”と表示する。文書名，作成者及び作成日付けは，実行時に利用者から指定された文字列とする。定義群は，この目的文書が目的言語処理系で実行中に，該当する定義の内容に置換される。
Helveticaは省略時想定書体名，12 _sは省略時想定文字寸法指定，<LM>及び<UM>はそれぞれ左余白，上余白を表わす正整数とする。座標原点は，媒体方向が肖像であるとして左下端とする。これらの省略時想定値は，文書本体の個別指定で変更できる。

a3．2．付せん区画の目的コード生成規則（marked area）

a3．2．1．書体名付せん及び文字寸法付せん（font & size）
書体名が指定されたとき，文字寸法が指定されたとき及び書体指定区画が終了したときには，次のコード生成を行なう。

```
_f1
<ポイント>   _sc
```

ここに，書体名は，Helveticaとする。書体名付せん区画が終了したときには，従前の書体名の区画に戻すものとする。 <ポイント>の省略時想定値は，10ポイントとする。

a3．3．原始文並びのコード生成規則（non－mark area）

a3．3．1．語（word）
原始文書の内，間隔と間隔で包囲された部分は，語とみなす。JIS漢字は，ここでは列10－15（GR）に呼び出す。

```
(  <文字符号列>  )   _s
   <文字送り量>  0      _rm
```

ここに<文字符号列>が，漢字符号のときJIS漢字表に各バイトの最上位ビットを1にした値を持つGR漢字符号とする。漢字符号の変換例を次に示す。

```
例：漢字の”漢”のJIS符号   =   03／4   04／1
    GR漢字符号             =   11／4   12／1
```

この符号を装置固有の符号に変換するのは，前置したプログラムによる。この漢字符号は操作系内部符号シフトジスにすることもできる。後者の場合には，前置する変換プログラムを，有賀（日本語PostSCript，卓上出版シンポジウム，情報処理学会，1988）のシフトジス変換を前置させることもできよう。

以上