

Title	FORTRAN77(V)における日本語文字データ処理
Author(s)	後藤, 米子
Citation	大阪大学大型計算機センターニュース. 1984, 52, p. 113-120
Version Type	VoR
URL	https://hdl.handle.net/11094/65601
rights	
Note	

Osaka University Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

Osaka University

FORTRAN77(V)における日本語文字データ処理

研究開発部 後藤米子

1. はじめに

昨年8月29日のFORTRAN77(V)コンパイラのバージョンアップに伴い、FORTRANのプログラムにおいても日本語の文字をデータとして処理することが可能になった。従来のFORTRANでは、1文字が1バイトで符号化されるものを処理できていたが、これに新規のデータの型として「日本語型」が設けられ、1文字が2バイトで符号化されるものも処理できるようになった。簡単に言えば、文字型のデータの扱いとほぼ同等のことが日本語型のデータについて行えること、印字文字の大きさを利用者が選択し指定できることなどである。これに伴い、入出力関係などに多少の機能が追加され、これらの新機能を有効とするためにコンパイラオプションJIPSが追加された。本稿で述べる新機能は、JIS規格にないことは言うまでもなく、FORTRANという言葉の性質からみると周辺的な拡張とみなせるので、当センターではこのオプションを運用の既定値に採用していない。

本稿では、2.で文字コードについて述べ、3.で文法の追加機能のうち宣言文および実行文(入出力文以外)について、4.で入出力関係について述べる。5.でファイルと書式つき記録の長さについて述べ、6.でページプリンタ^{注1)}に出力するための書式制御情報名とそのジョブ制御言語(以下、JCLと略す)などについて述べる。日本語文字データの作成方法については本稿では述べないので、参考文献^{6,7)}を参照していただきたい。

2. JIS規格とACOS-6のコード系の比較

JIS規格¹⁾では、(a)情報交換用符号系 JIS C6220(1976) および(b)情報交換用漢字符号系 JIS C6226(1978)により、2種類のコード系が規定されている。JIS C6220は1文字が1バイトで符号化されるものについて、JIS C6226は1文字が2バイトで符号化されるものについての規格である。ACOSシステム1000では、上記に対応したコードをそれぞれJISコードおよびJIPS^{注2)}コード^{7,11)}と呼び、これらのコードで表現される文字をJIS文字およびJIPS文字と呼ぶ。JIS C6226規格で規定されているコードは、ACOSシステム1000では同一のコードで使用できるが、規定外のものについてはACOSシステム固有のコードを使用することになる。

注1) 日本語プリンタまたはNIP(Non-Impact Printer)とも呼ばれる。

注2) JIPSは、Japanese Information Processing Systemの略である。

表1 JIS C6226(1978)とJIPS文字の字種

文字種区分	定義文字数	備考
記号	108	JIS C6226に 規定
英数字	62	
平仮名	83	
片仮名	86	
ギリシア文字	48	
ロシア文字	66	
JIS漢字第1水準	2,965	
JIS漢字第2水準	3,384	
添字(英数字・記号)	94	JIPS固有
添字(片仮名)	94	
その他(罫線など)	471	
合計	7,461	

JIS C6226とJIPS文字の字種数の比較を表1¹¹⁾に示す。JIPS文字は、規格に罫線および添字用の文字などの合計659が追加されている。本稿ではJIPS文字の集合全体を日本語文字と呼ぶが、このなかには漢字のみならず、ギリシア文字および学術用の特殊記号なども含まれていることに注意を要する。

3. FORTRAN 77(V)の日本語処理機能(1) — 入出力関係以外

本節および4.で述べる新規追加機能は、すべてコンパイラオプション⁴⁾JIPSを指定したときに使用可能となるため、利用者は、バッチ処理では\$FRT77文で、TSS処理ではRUNコマンドでこのオプションを必ず指定しなければならない。下記には主要な項目について示す。特に断らない項目については、文字型²⁾に準じて取扱われる。

3.1 日本語型データ

日本語型データは、JIPS文字によって構成された列であり、日本語定数では1~128文字、これ以外では1~256文字から成る。一つのJIPS文字は2バイト(18ビット)を占める。

日本語定数には次のものがある。 n はJIPSコードを示す4桁の16進数、 l はJIPS文字、 h はJIS文字である。 h はそれぞれ対応するJIPS文字に変換されるが、直後が濁点または半濁点の文字はその2文字が1つのJIPS文字となる。(3)では片仮名が平仮名に変換される。

- (1) $NX^{\nabla} nn \cdots n^{\nabla}$
- (2) $NC^{\nabla} ll \cdots l^{\nabla}$
- (3) $NH^{\nabla} hh \cdots h^{\nabla}$
- (4) $NK^{\nabla} hh \cdots h^{\nabla}$

上記以外に索引定数があるが、索引辞書を指定する必要があるため省略する。日本語端末が手元にないときは主として(1)の形式を使用する。例えば文字列 $\alpha\beta\gamma$ は $NX^{\nabla}264126422643^{\nabla}$ となる。

3.2 宣言文およびDATA文

- (1) 日本語型の型宣言文は、NCHARACTER文であり、次のことを除いてCHARACTER文に同じである。
 - 長さ指定に整合文字長指定は許されない。
 - 文字列の長さは1～256文字である。
- (2) 一つの共通ブロック内に、日本語型と他の型が混在してはならない。
- (3) EQUIVALENCE文で、日本語型のデータを他の型のデータと結合してはならない。
- (4) DATA文で日本語型の配列名に初期値を設定する場合は、初期値が設定される要素と初期値を表す定数または定数名の間には必ず一対一の対応がなければならない。

3.3 実行文

- (1) 演算子 // (日本語演算子と呼ぶ)により、日本語型データのみを互いに連結することができ、その形式は文字演算子と同じである。日本語演算子の優先順位は、文字演算子に等しい。
- (2) 日本語型データに空白が補われる場合は、2バイトの空白が使用される。
- (3) 次の組込み関数が追加され、これに伴い総称名⁴⁾が追加された。

組込み関数	引数の個数	総称名	個別名	引数の型	関数の型
長さ	1	LEN	LEN	文字型	4バイト
			NLEN	日本語型	
部分列の指標づけ	2	INDEX	INDEX	文字型	整数型
			NINDEX	日本語型	

4. FORTRAN 77 (V) の日本語処理機能 (2) — 入出力関係

4.1 入出力文

OPEN文およびINQUIRE文の指定子の並びに、SHIFTCODE = sc が追加された。

- OPEN文では sc は文字式であり、後続する空白を除いた値が、APPLIED (シフトコード付加、既定値) または SUPPRESSED (シフトコードなし) である。書式つきで接続するときに指定でき、この性質は出力の場合にのみ意味をもつ。
- INQUIRE文では sc は文字型の変数名または配列要素名であり、シフトコードの有無に応じて前述の値が返却される。

4.2 編集記述子

編集記述子として次のものが使用できる。 w および w' は欄の中を示し、 $w' = 2w$ の関係がある。

ただし、 $1 \leq w \leq 256$ 。

反復可能編集記述子 N, Nw, Gw, A, Aw' 。(日本語型の入出力並び項目と対応)

反復不能編集記述子 $\left\{ \begin{array}{l} NX^\nabla nn \cdots n^\nabla, NC^\nabla ll \cdots l^\nabla, NH^\nabla hh \cdots h^\nabla, NK^\nabla hh \cdots h^\nabla. \\ NP_s, NT_c \text{ (ページプリンタ出力にのみ使用)}. \end{array} \right.$

- N, Nw および Gw は、文字型の編集記述子と同じ動作をする。
- A および Aw 編集記述子で出力するとき、シフトコードは付加されないで、これらをページプリンタ出力に使用できない(5.1参照)。
- $N, Nw, Gw, NX, NC, NH, NK$ 編集記述子で出力するとき、欄の中の前後にシフトコードが付加される。ただし、OPEN文のSHIFTCODE指定子の値がSUPPRESSEDの場合には、付加されない。
- NP_s 編集記述子は、ページプリンタに出力する場合に、文字の大きさ s を各書式制御情報名(6.1参照)の既定値以外に変更するとき使用する。 s はポイント数であり、次のいずれかの値である。

s の値 7, 9, 12, 14, 18, 24.

一つの書式内において、次の NP_s 編集記述子に出会うまで有効である。既定値より大きい s の値を指定するときには、この印字行の直前に1行の空行がなければならない。

- NT_c 編集記述子は、ページプリンタに出力する場合に、出力桁の位置決めに使用する。印刷行の左端を第1桁として c 桁目が印字開始位置となり、このときの1桁は10CPI(6.1参照)の文字間隔で書式制御情報名にかかわらず固定である。一つの書式内で、これと TR_c, TL_c および T_c 編集記述子とを併用してはならない。すでに記録が出力されている桁に後退して出力してはならない。

5. JIPSコードを含むファイルと記録

5.1 シフトコード

FORTRAN77(V)における標準のファイル形式は、JIPSコードの記録を含む場合も順編成および直接編成にかかわらず、UFAS(V)のFRC形式である。一つの記録は、JIPSコードとJISコードの混在で構成されてもよい。書式つき記録においては、システムの標準では両者の間は2バイトのシフトコードで識別され、JISコードからJIPSコードへの切替えを“漢字イン”(KIと略す)で行い、この逆を“漢字アウト”(KOと略す)で行う。各記録の先頭ではJISコードが仮定されるので、行送り文字を含まないJIPSコードのみからなる記録でも各記録の最初と最後に、それぞれ“KI”と“KO”が付加される。図1にその例を示す。シフトコードは、ファイルをページプリンタまたは漢字端末に印字または表示する場合に必要となるもので、システムの標準では自動

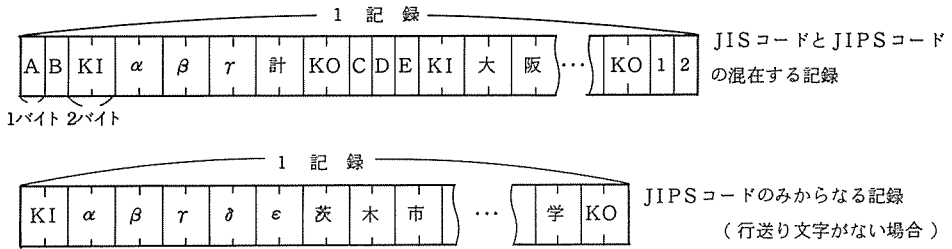


図1 JIPSコードを含む書式つき記録の例

的に付加されるが、プログラム内でのみファイル进行处理する場合には、付加されていなくてもよい。書式なしファイルにおいては、シフトコードは付加されない。

5.2 記録の長さ

JIPSコードを含む書式つき記録の長さは、次のようになる。これは、JIPS文字の一つが2¹¹⁾バイトを占めること、および各種の機能文字を含むためである。

$$\text{記録の長さ (バイト)} = (\text{JIS文字の数}) + (\text{JIPS文字の数}) \times 2 + (\text{機能文字の占めるバイト数})$$

FORTRAN 77 (V)では、順編成書式つき記録の長さの標準値は133⁵⁾バイトである。この値を拡張変更するには次のいずれかを使用する。特にページプリンタに出力する場合には、その値として300 (バイト)を指定すると十分である。

- OPEN文のRECL指定子。^{3,5)}
- 組込みサブルーチンLONGIO。⁵⁾
- { \$DDEF文 (バッチ処理)。⁵⁾
 { RUNコマンドのDDEFオプション (TSS処理)。⁵⁾

\$DDEF文では、ページプリンタを表す装置番号は“P*”となることに注意されたい。

JIPSコードを含む書式なし記録の長さ⁵⁾は、日本語型の各出力並び項目についてその長さの2倍となること以外は従来と同じである。なお、FORTRAN 77 (V)における記録の長さについては、参考文献^{5,8)}を参照していただきたい。

6. ページプリンタへの出力

6.1 書式制御情報名

ページプリンタがラインプリンタと異なる点は、利用者が印字文字の大きさ、字体および1ページの印字行数などを指定できることである。この情報を書式制御情報と呼ぶ。ACOSシステム1000では、JOSPと呼ばれるユーティリティプログラムを使用して、利用者がこの指定を行うようになっているが、本センターでは、表2に示すような標準的なものをシステムに登録し提供している。利用者は6.2で述べる方法により、ページプリンタに出力できる。以下では、書式制御情報名とは

表2 書式制御情報名(センター提供)一覧

書式制御情報名	行密度		JIPS文字			JIS文字		
	LPI	行/頁	文字の大きさ	CPI	字/行	文字の大きさ	CPI	字/行
NORMAL	6	62	7P	10	140	7P	10	140
NORMAL 9			9P	8	112			
NORMAL 12	4	41	12P	6	84			
NORMAL 14			14P	5	70			
NORMAL 18			18P	4	56			
NORMAL 24	3	31	24P	3	42			
GOTHIC 10	6	62	9P	10	140	7P	10	140
GOTHIC 12	8	82	7P	10	140		12	168
GOTHIC 15			なし				15	210

LPI : Line per Inch
CPI : Character per Inch

表2のものを示す。

各書式制御情報名の行密度、文字の大きさおよび文字間隔の既定値は次のとおりである。

- 行密度の単位はLPI (Line per Inch)で、この値から算出した1ページあたりの印字行数を示す。
- 文字間隔の単位はCPI(Character per Inch)で、この値から算出した1行あたりの印字文字数を示す。

NP_s編集記述子(4.2参照)で既定値の文字の大きさを変更指定したとき、その文字の大きさに応じた文字間隔がとられる。

書式制御情報名のNORMAL系およびGOTHIC系の特徴を次に示す。NORMAL系はJIPS文字からなるものに使用し、GOTHIC系はJIS文字からなる英文の清書などに使用する。

(1) NORMAL系(6種類)

- JIPS文字の標準のポイント数を次のなかから選択できる。

7, 9, 12, 14, 18, 24.

- JIS文字の大きさはつねに7ポイントで、その字体は、英大文字はラインプリンタと同一(OCR-Bフォント)、英小文字はゴシック体である。

(2) GOTHIC系(3種類)

- JIS文字の大きさはつねに7ポイントで、その字体は英大文字および英小文字ともにゴシック体である。
- JIPS文字の大きさは7ポイントまたは9ポイントを指定でき、その字体は(1)と同一であ

る。

6.2 ジョブ制御言語およびTSSコマンド

ページプリンタに出力するためのJCLおよびTSSコマンドは次のようになる。

(1) バッチ処理

\$JOB文に出力先を示すパラメータJPRを指定し、かつ\$JOBDEF文で書式制御情報名を一つ指定する。\$JOBDEF文の標準的な位置は\$JOB文の次である。[]は省略できることを示す。

```
1      8      16
$      JOB      課題番号 $パスワード, ジョブクラス, [識別名], [R], [端末識別名], JPR
$      JOBDEF   FORMSET = 書式制御情報名
```

(2) TSS処理

すでに作成されているファイルに対して、\$FPRINTJコマンド(省略形は\$FPRIJ)を入力すると、

```
FORMSET?
```

の問合せがあるので、その応答として書式制御情報名一つを入力する。

6.3 注意事項

ページプリンタでは、ラインプリンタと比べて次のことに注意する必要がある。

- 重ね打ちを使用できない。
- 用紙1枚ごとに上下に印字不能域があるため、ミシン目をはさんで空白が生じる。

ラインプリンタとほぼ同一の文字の大きさおよび行間隔を得たいときには、書式制御情報名には“NORMAL”を指定するとよい。

7. おわりに

ACOSシステム1000の日本語処理は他機種に比べ遅れていたが、本センターにおいてもようやく昨年4月頃から少しずつではあるが実現されつつある。本稿で述べたことは、日本語処理全体からみるとごく限られた一端であり、FORTRANプログラムで何ができるかということを紹介したにすぎない。これらの機能を活用した簡単な一例として、計算結果の見出しなどに漢字やギリシア文字などを印字することが考えられるが、その手引として本稿が役立つは幸いである。

本稿を作成するにあたり、有益な助言をいただきました本センター研究開発部大中幸三郎助手に感謝いたします。

参 考 文 献

- 1) JISハンドブック情報処理 1983, pp.231-240, pp.264-323, 日本規格協会(1983).
- 2) 大中, 後藤: FORTRAN77(V)概説(1), 大阪大学大型計算機センターニュース, Vol.11, No.3(1981).
- 3) 大中, 後藤: FORTRAN77(V)概説(2), 大阪大学大型計算機センターニュース, Vol.11, No.4(1982).
- 4) 大中, 後藤: FORTRAN77(V)概説(3), 大阪大学大型計算機センターニュース, Vol.12, No.1(1982).
- 5) 後藤, 大中: FORTRAN入出力文からみた記録の長さ, 大阪大学大型計算機センターニュース, Vol.13, No.2(1983).
- 6) 中島: 日本語エディタ(NEDIT), 大阪大学大型計算機センターニュース, Vol.13, No.3(1983).
- 7) 北本, 他: 漢字端末について, 大阪大学大型計算機センターニュース, Vol.13, No.3(1983).
- 8) FORTRAN77(V)とFORTRAN(R)のバージョンアップ, 大阪大学大型計算機センター速報, No.107 (1983).
- 9) AGB 01-1 FORTRAN77 言語説明書, 日本電気(1984).
- 10) FGB 09-2 FORTRAN77(V)プログラミング手引書, 日本電気(1984). (印刷中)
- 11) FKZ 01-3 日本語情報処理システム概説書, 日本電気(1983).
- 12) FKA 03-1 サービスプログラム説明書, 日本電気(1983).
- 13) FKC 01-2 日本語エディタ説明書, 日本電気(1983).