

SNOBOLインタプリタ・システムの評価

緒方博通 豊田順一 田中幸吉
(大阪大学 基礎工学部)

1. はじめに

現在においては、計算機システムの利用は数値計算のみならず文献検索等の記号処理まで広がって来た。ところが一般に記号処理システムの記号処理に要する時間は長いといわれている。この原因として、高位の計算機でなければ機械語レベルの高度の記号処理命令がないことや、対象問題自体に多くの処理プロセスが含まれていることが挙げられる。このような背景にもとづき、ハードウェア化に適する記号処理命令はどうあるべきかの立場より先にSNOBOLインタプリタを開発した⁽¹⁾⁽²⁾。そして更に共通の記号処理概念で、プログラムの前処理を行なうなどしてモッサリアップをめざし同時に機能の拡大を行なった⁽²⁾⁽³⁾。そして、本報告ではその両者のインタプリタに対する評価を行なうものである。

2. 評価の対象

両者のインタプリタはミニコンCEC555H (PDP8相当)に構成されている。詳細は参考文献⁽¹⁾⁽²⁾⁽³⁾にあるので省略する。両者をそれぞれ、使用する補助記憶からCassette SNOBOLインタプリタ(以下Cassette SNOBOLと略す)とDisk SNOBOLインタプリタ(以下Disk SNOBOLと略す)と呼ぶ。両者の主たる相違点を表2.1に示す。Disk SNOBOLでは実行前に、ラベル名、定数、変数名等を表にすると同時に、ソースプログラムからコード化したプログラムを作成して、実行時にスキャンする文字数を減らすなどして、実行時間の短縮を目指している。

両者のインタプリタでの記号処理は、すべて記号処理演算領域で行なわれる。すなわち、必要なデータのコピーを演算領域に搬送して、処理結果をデータ領域に戻す方式である。すなわち、使用メモリ空間を記号処理演算領域とデータ領域とに明確に分離し、記号処理過程を概念的に記号処理演算過程とデータ更新過程に分離などしている。Cassette SNOBOLとDisk SNOBOLでは、共通の記号処理概念で記号処理を行なっているが、細部が異なるので以下Disk SNOBOLについて説明する。

記号処理演算領域は表2.2のような4個の記号処理レジスタと呼ばれるものより構成される。ただし、ハードウ

表2.1 両インタプリタの比較

相違点	Cassette SNOBOL	Disk SNOBOL
使用コア	12 K語	24 K語
補助記憶	カセット	ディスク
実行前の処理	表作成	表作成 ソースプログラムのコード化
表引き	シーケンシャルサーチ	アドレス・アサインメント
言語水準	ミニSNOBOL ⁽¹⁾⁽²⁾	ミニSNOBOLに追加機能 ⁽²⁾⁽³⁾

表2.2 記号処理レジスタ

レジスタI (RI)	新記号列生成
レジスタII (RII)	対象記号列格納
レジスタIII (RIII)	パターン記号列格納
レジスタIV (RIIV)	選択記号列格納

レジスタが用意されていないので、ソフトウェアで代用する。レジスタ内の中間結果や最終結果は、部記号列表示用の指示子で示される。指示子の値はレジスタ内の絶対番地である。

レジスタ間の記号処理演算は記号処理過程を基本操作に分解し、かつ統合可能な操作をまとめた結果、照合演算と転送演算とした。図2.1はバランス照合の例を示している。(2)は変数V1の

RI
RII
RIII

ストリング中の"0"と"L"とでバランスしているストリングを変数N1として、V1は照合が成功した場合にはその部分を"??"で置き換える SNOBOLステートメントである。バランスしているとは、この場合では"0"と"L"の対応がついているということである。(1)、(2)の後でV1とN1の値はそれぞれ、"SN??ANGUAGUE"と"BOLE"となる。

V1 = "SNOBOL LANGUAGE" (1)
V1 "0" * (N1) * "L" = "??" (2)

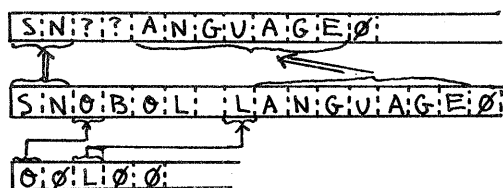


図2.1 バランス照合の例

次に変数値の格納の仕方について述べる。すべての変数の値はデータ領域と呼ばれる領域に許される限り連続に格納されるが、既使用セルに遭遇するとポインタにより、次の未使用セルへとリンクされ格納される。図2.2に示す。新しい変数値の登録は以下の通りである。

ステップ1

変数表を参照して、データ領域へのポインタを得る。

ステップ2

古い変数値があるかテストする。なければステップ4へ進む。

ステップ3

ステップ1で得たポインタによって古い変数値を消去する。

ステップ4

新しい変数値を格納するデータ領域の場所を確保する。

ステップ5

新しい変数値をRIよりデータ領域へ転送する。また表更新を行なう。

また、ゴミ集めはデータ領域内で使用した最大番地を記録しておき、その番地がデータ領域にセットされた番地を越え

そうになった時、行なう。

その他の処理プロセスとしては、SNOBOLソースプログラムのコード化処理、コード化されたプログラムのオーバレイ処理、整数型四則演算、組込み関数、サブルーチン、間接ラベル、間接変数、配列等の処理があるが、ここでは略す。

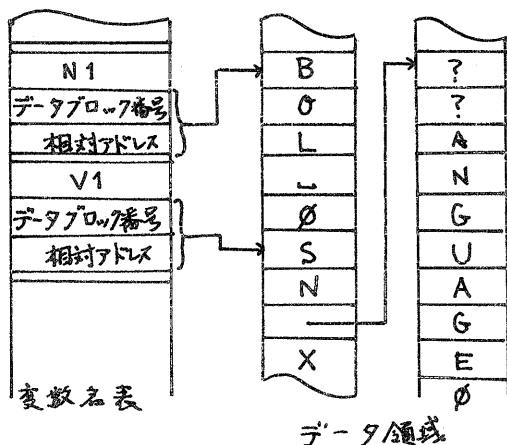


図2.2 データの格納

3. インタプリタ評価の方法

本インタプリタの評価は、各種記号処理操作と記号処理ステートメントのデ-

タ長による処理時間の变化と、各種記号処理プログラムの処理時間の解析によって行なった。

ところで、本インタプリタの構成されているCEC 555 Hには処理時間の計測のためのハードウェアは用意されていないので、入出力命令によって駆動する簡単なハードウェアを作成した⁽⁴⁾。入出力命令をそれぞれ帰零、計時開始、計時中断に対応させる。例えば、帰零信号はCEC-555 Hでは図3.1のようにして得られる。そして、入出力命令を計測すべきルーチンの入口と出口に挿入する。すなわち、入口に計時開始、出口に計時中断に対応する入出力命令を挿入するわけである。そして、その入出力命令によって生じた信号と水晶発振器によって得られたクロックパルスにより、そのルーチンの使用回数を示すカウンタと処理時間を示すクロックを駆動する。図3.2にこれを示す。クロックパルスは1MHzで、クロックは最小単位1μ秒で10進8桁、カウンタは10進4桁であり、1度に4個のルーチンの計測が可能のように、4組のカウントとクロックを用意した。図3.3に全容を示す。

この装置により、処理時間の実時間計測を行なうことができ有用であった。なお、入出力命令挿入による増分は補正した。

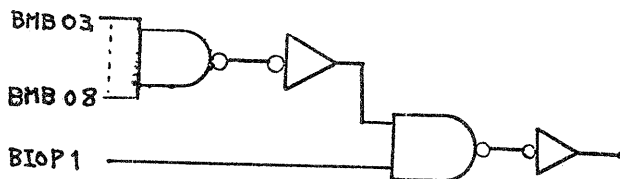


図 3.1 帰零信号

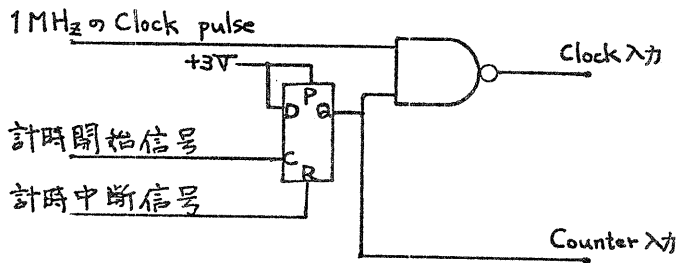


図 3.2 カウンタとクロックの入力信号

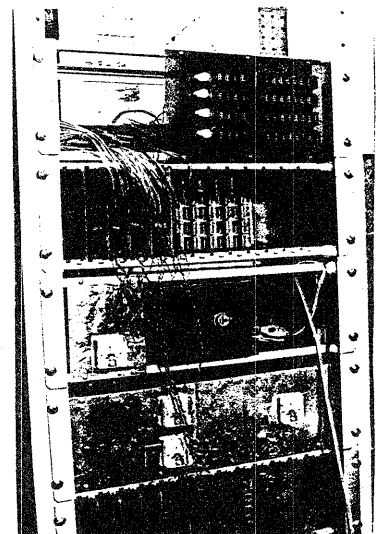


図 3.3 処理時間計測装置

4 インタプリタの評価

4.1 Cassette SNOBOLとDisk SNOBOLとの比較

2に述べた通りCassette SNOBOLとDisk SNOBOLには種々の相違点があるので、記憶媒体の違いを除いた処理時間の相違を計測した⁽⁵⁾⁽⁶⁾。両者で共通に実行できるSNOBOLプログラムとして三者を選び、計測した。それは各々、次のようなプログラムである。右端に実行SNOBOLステートメント数を示す。

- | | | |
|------|---------------------------------|------|
| No.1 | 置換によって図形を生成するプログラム | 5073 |
| No.2 | tree searchを行なうプログラム(11のプログラム例) | 2747 |
| No.3 | 入力データ中の決った記号列を他の記号列に置換 | 1942 |

これによって、表4.1 っような結果を得た。入出力及び補助記憶による相違を差し引いた比を最下段に示す。

次に実行ステートメント数を変えないように留意しながら、同一プログラムッデータ長を変えて処理時間を計測した。入出力時間と補助記憶との交信時間を除く両者の処理時間とその比を表4.2に示す。

表4.2 両インタプリタッ比較
同プログラムで
(単位 msec) データ長を変えて
I/O 時間を除く

データ長	1	2	3	4
Cassette ⁽¹⁾	380666.0	413453.0	445189.6	447645.5
Disk ⁽²⁾	16031.7	20465.8	25169.8	28764.6
(1)/(2)	23.7	19.7	17.3	16.3

両計測により、Disk SNOBOLはCassette SNOBOLと比して6倍以上の向上となっているが、明らかにになった。これは、表引き方式の変更と、プログラムッコード化によるものである。また、対象問題やデータ長により、両者の比の異なるのが明らかである。これは、データ長が大になるにつれ、処理に要するプロセスが多くなると、全体の処理時間が増大し、プログラムッコード化と表引き方式の変更によるスピードアップへの寄与が減少するためである。

4.2 Disk SNOBOLの記号処理操作時間の評価

4.2.1 計測

両インタプリタは基本的な記号処理の概念はほぼ同一である。そこで以下ではこの記号処理の概念が妥当であるかをDisk SNOBOLの評価を通して明らかにして行く。

まず、Disk SNOBOLの基本的な5つの記号処理操作がデータ長によってどのように処理時間を要するかを計測した。

(1) ロード操作

ロード操作によって、データ領域から記号処理レジスタへ、ストリングが転送される。ストリング長と処理時間との関係を図4.1に示す。処理時間 T_L とデータ長 L_L との関係式は

$$T_L \approx 0.07 * L_L + 0.2 \quad (\text{msec}) \quad (4.1)$$

である。

(2) ストア操作

記号処理レジスタからデータ領域へストリングを転送する操作である。この処理時間の様子を図4.2に示す。処理時間 T_S とストリング長 L_S との間の関係式は

表4.1 全処理時間の比較

program name items		No.1 HEART	No.2 PASS	No.3 CHANGE
program size statement		35	67	13
CASSETTE SNOBOL	C1) total processing time	1034983.5 (msec)	435386.1 (msec)	383563.5 (msec)
	C2) cassette active time	31723.8	53642.6	12920.9
	C3) input-output time	1964.4	1077.5	24516.3
	C4) C1 - C2 - C3 time	1001295.3 (msec)	380666.0 (msec)	346126.3 (msec)
DISK SNOBOL	D1) total processing time	151998.3 (msec)	17152.6 (msec)	69380.3 (msec)
	D2) input-output time	1231.6	1120.9	58437.1
	D3) D1 - D2 time	150766.7 (msec)	16031.7 (msec)	10943.2 (msec)
	C4 / D3	6.6	23.7	31.6

(4.2) のように求められた。

$$T_s \approx 0.09 * L_s + 0.2 \quad (\text{msec}) \quad (4.2)$$

(3) 転送操作

記号処理レジスタ間で、サブストリングを転送する操作である。図4.3に示す処理時間 T_t とストリング長 L_t との関係式は以下のようになる。

$$T_t \approx 0.07 * L_t + 0.2 \quad (\text{msec}) \quad (4.3)$$

(4) 消去操作

本操作によって更新されたストリングの古い値をデータ領域より消去する。処理時間を T_c 、古い値のストリング長 L_c との関係式は(4.4)のように求められた。

$$T_c \approx 0.06 * L_c + 0.2 \quad (\text{msec}) \quad (4.4)$$

(5) 照合操作

照合操作は、与えられたストリングがあるサブストリングを含むか否かを判定するものである。この操作の処理時間にはいくつかの要素がある。例えばストリング長や、サブストリングを含むか否か、照合操作の深さ等である。そこで、照合の成功する例としてストリング長が同一で、照合操作処理時間が最大、最小となる場合を取り上げた。

- i) 最大の場合 $A^{m-1}B$ と $A^{n-1}B$ ($m \geq n$) の照合
- ii) 最小の場合 A^m と A^n ($m \geq n$) の照合

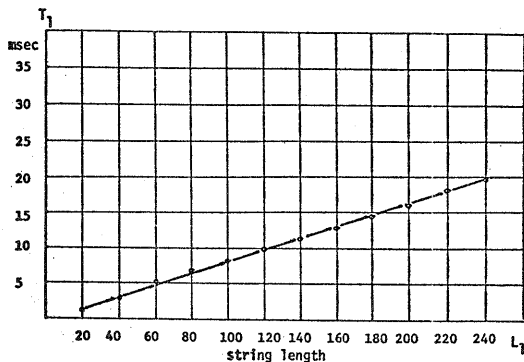


図4.1 ロード操作処理時間

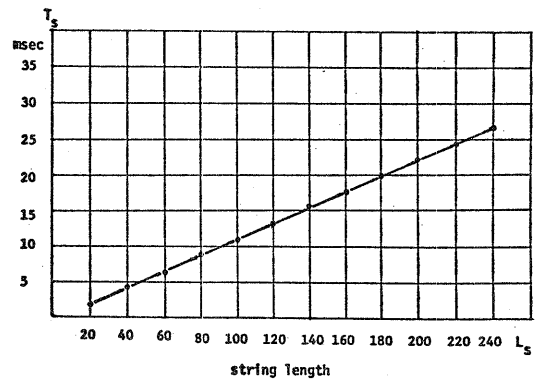


図4.2 ストア操作処理時間

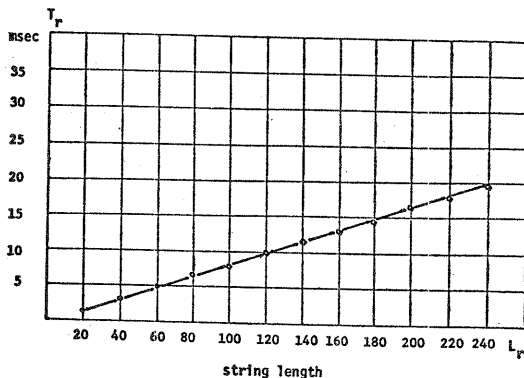


図4.3 転送操作処理時間

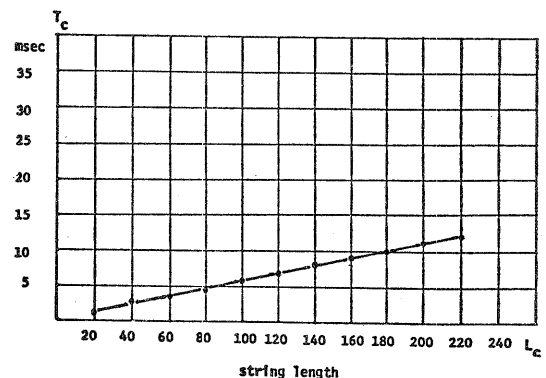


図4.4 消去操作処理時間

それぞれ図4.5, 図4.6に処理時間を示す. i), ii) の文字の照合回数は各々, $(m-n+1)n$ と n である. 照合に失敗して, ポインタを元に戻す回数はそれぞれ, $(m-n)$ と 0 である. 図4.5, 図4.6より, 照合回数を L_m , 失敗回数を L_f とすれば, 照合操作処理時間 T_m は

$$T_m \approx 0.03 * L_m + 0.06 * L_f + 0.3 \quad (4.5)$$

(msec)

で示される.

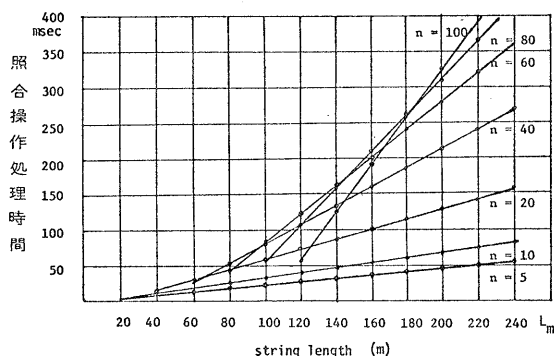


図4.5 $A^{m-1}B$ $A^{n-1}B$ 照合操作処理時間

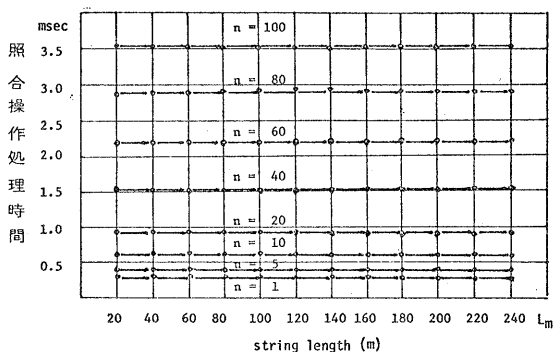


図4.6 $A^{m-1}A$ $A^{n-1}A$ 照合操作処理時間

4.2.2 考察

4.2.1 で各操作の処理時間を明らかにした. これより, 実験値は線形性を示している. これは, 各操作がストリングの内容に関係なく, 単純にその操作を繰返すためである. そのため, 処理時間は処理するストリングの長さが大きくなるに従って増大していく. 一回当りの処理時間は $0.03 \sim 0.06$ msec と少ないが, 処理回数が多いため, 全体の処理時間は長くなってしまふ. 一回当りの処理時間 $0.03 \sim 0.06$ msec も高次の記号処理命令を有す計算機と比して長い. 特にロード, ストア操作の処理は時間がかかっている. この一回当たりの処理時間の大きさが全体の処理時間を制していることが予想される.

4.3 各 SNOBOL ステートメントの処理時間

4.3.1 計測

SNOBOL の各記号処理ステートメントの処理時間は4.2 に述べた操作処理時間とその文の解釈, 表引き等の時間より成り, 多くの要素がある. 例えば, SNOBOL ステートメント特有の並び先指定があれば, その処理のために処理時間が増大する. それらの解釈等の時間はステートメントの形などによって異なる, そこでステートメントのいくつかについて, 処理時間を計測した.

(1) 代入文

代入文の処理時間は大きく分けて次の4つより成る.

① 解釈や次のステートメントへの移動等ステートメントの形に依存する処理時間

② 4.2 の (1), (2) の記号処理レジスタ領域とデータ領域間の搬送処理時間

③ 4.2 の (4), すなわちすでに変数に値が与えられている際, その値

の解放時間

④変数値を格納する際のデータ領域の空き探索時間

①はステートメントで使われる変数やラベルが間接変数やラベル変数でなければ約1.4 ~ 1.8 msecである。②, ③は4.2で述べた通りである。④はデータ領域の状態による値であるので測定より除外した。③, ④を除いた処理時間を図4.7に示す。

(2) 連結文

(1)の代入文と同様に①~④より成る。同様に図4.8に示す。

(3) 照合文

照合は記号処理レジスタにstring reference と pattern を転送してから行なわれるので②を伴なう。他に

⑤ 4.2.9 (5) すなわち記号処理レジスタ上での照合操作処理時間が含まれる。値の更新は伴わないので③, ④は含まれない。⑤はstring reference と pattern の長さにより異なる。4.2と同様に処理時間が最大、最小となる例として A^mB と $A^{n-1}B$, A^m と A^n について扱った。図4.9, 図4.10に処理時間を示す。

(4) 置換文

置換は成功する場合と失敗する場合で、処理時間がかなり異なる。失敗する場合は照合ステートメントの処理時間とはほぼ同処理時間である。成功する場合に

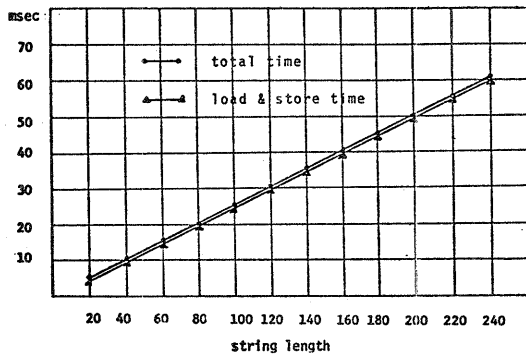


図4.7 代入文処理時間

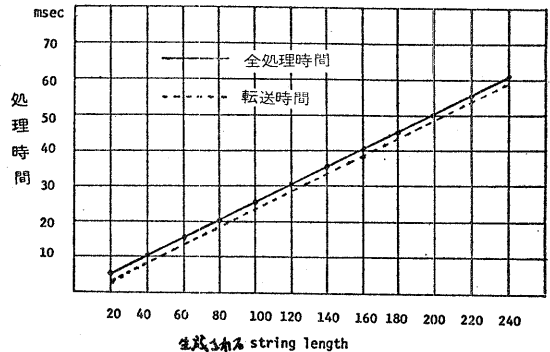


図4.8 連結文処理時間

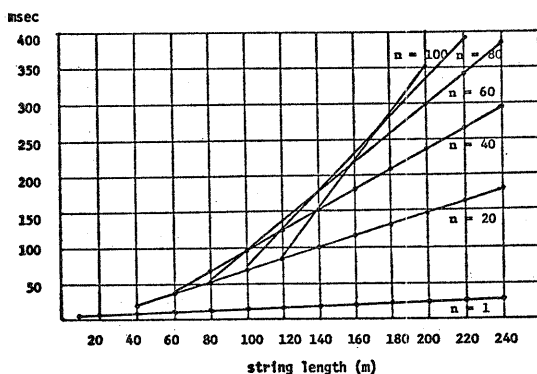


図4.9 $A^{m-1}B$ $A^{n-1}B$ 照合文処理時間

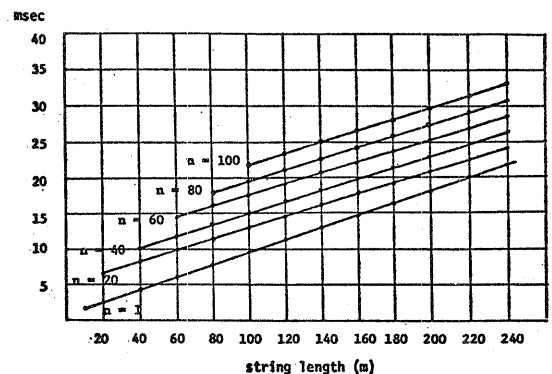


図4.10 $A^{m-1}A$ $A^{n-1}A$ 照合文処理時間

は値の更新を伴うので、処理時間は①～⑤より成る。照合文と同様に成功する際の処理時間が最大、最小となる場合について計測した。④を除く処理時間を図4.11、図4.12に示す。

(5) その他

その他の記号処理ステートメントとして、バランス照合、固定長変数、ストリング変数などが掲げられるが、(4)と同様な性質を示す。

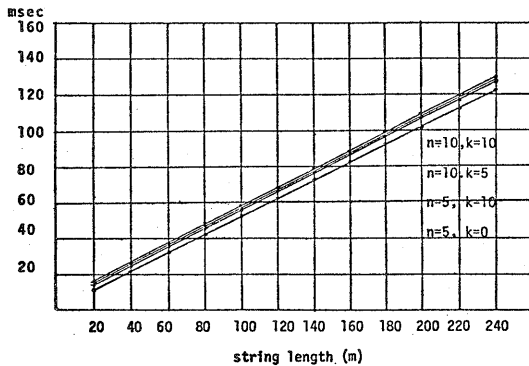


図4.11 置換文処理時間

成功する場合で最大処理時間

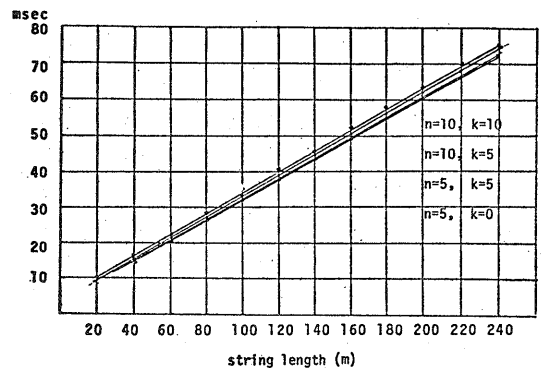


図4.12 置換文処理時間

成功する場合で最小処理時間

4.3.2 考察

4.2と同様に線形性を示すが、これは4.2をもとに実現されているためである。各SNOBOLステートメントの処理時間の内、特に代入文、連結文では②の転送処理の占める割合が多い。特にデータ長が大になるにつれて処理プロセスが増大し、処理時間に占める解釈時間の占める割合が減少して、②の転送の処理時間の占める割合が増大する。照合文、置換文等では照合するデータにもよるが、やはり転送処理時間の占める割合は大きい。また①の解釈時間も無視できない。①は処理の多少にかかわらず、1ステートメント当り約1～2 msecかかるわけで、この減少が望ましい。

4.4 各種記号処理プログラムにおける処理時間

4.4.1 計測

前節までに、各SNOBOLステートメントの処理時間と各記号処理操作の処理時間を明らかにした。本節では、SNOBOLで書かれた種々の記号処理プログラムを実際に実行させてみて処理時間を計測した。

まず、Disk SNOBOLのコード化の処理時間とソースプログラム中のステートメント数との相関を示したものが図4.13である。この内の80～90%はdiskとの交信時間である。処理時間はかなり大きいプログラムでも急激な処理時間の増大はない。

次に、コード化された各種の記号処理プログラムを実際に実行して、処理時間を計測したが表4.1である。対象問題によって、かなりのばらつきが見られる。それぞれ例は以下の通りである。

例1 アセンブラのサブルーチン名を求め、その間の関係を求める。

- 例2 SNOBOL 4 の組込み関数を SNOBOL3
で実現のためトランスレータ
例3 文献検索の内主題分析
例4 4.1 の No.1 と同一
例5 文献検索
例6 4.1 の No.2 と同一
例7 FORTRAN の論理 IF を算術 IF に変
換するトランスレータ
例8 4.1 の No.3 と同一
例9 LSP (FORTRAN) トランスレータ⁽⁷⁾

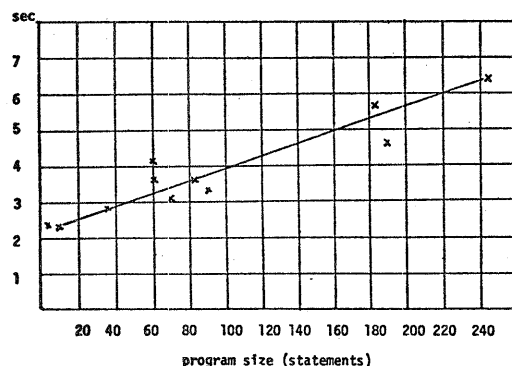


図4.13 前処理時間

4.4.2 考察

表4.1に示されたように解釈時間

と転送処理時間がかかりの部分を占めていることが明らかになった。解釈時間は表引きの回数とコード化されたソースプログラムを読んだ回数が多いことが処理時間を増大させる要因と思われる。転送処理時間が大きな割合を占めるのは、4.2, 4.3 で予想される通り、記号処理が記号処理レジスタで統一的行なわれるためである。当初予想された照合操作の処理時間は比較的小さな割合しか占めていない。これは他の処理時間が大きな割合を占めたためである。

5. おさげ

以上のように両者のインタプリタの比較、基本的記号処理操作、記号処理スタートメント、各種記号処理プログラムの処理時間の計測を行なった。

その結果、データ処理を記号処理演算領域で行なうため、データ領域と記号処理演算領域間の転送処理時間が大きいことが明らかになった。特にデータ長が大になるにつれてこの傾向は強い。この時間を減少させるには、処理のハードウェア化によって、転送を一文字単位ではなくて、ブロック単位にする等の対策が考えられる。

また解釈時間も無視できない。これはインタプリタ方式を取ったことに原因の大半がある。SNOBOLにおいては実行時に決定される要素が多く、これがコンパイラ方式を妨げる原因である。しかし、コード化の処理時間に占める割合が少ないこともあり、もう少し前処理を補助させる必要がありそうである。

照合操作処理時間の占める割合は小さいが、これは他の処理プロセスの占める割合が大きいため、他の処理プロセスの高速化が実現されると早晩問題となろう。その高速化にも連想記憶等のハードウェアが必要であらう。

最後に、本研究にあたり終始議論に参加いただき有効な助言を多数いただいた田中研究室の各位、とりわけ水本朋晴助手、三上和敬氏(現三菱電機)、馬野元香氏、宮下厚氏に深謝の意を表わします。

参考文献

- (1) 三上, 豊田, 田中: "新しい記号処理概念による SNOBOL インタプリタ" 情報処理 Vol. 16, No. 6 (1975)
- (2) 三上, 豊田, 田中: "コンパイラインタプリタ方式による SNOBOL 処理システムの構成" 情報処理学会第16回大会予稿集
- (3) K. Mikami: "SNOBOL INTERPRETERS ON A MINI-

表 4.1 各種記号処理プログラムにおける処理時間 (6)~(13)は処理時間×処理回数
(18)に占める割合(%) 平均処理時間

	例1	例2	例3	例4	例5	例6	例7	例8	例9
(1) 実行時間	413817	14376	12943	5073	4735	2747	2515	1942	1399
(2) 表引き回数	114149	17635	22306	10018	8443	7223	3695	2622	950
(3) ②/①	0.3	1.2	1.7	2.0	1.8	2.6	1.5	1.4	0.7
(4) コードを読込む回数	476677	146787	121876	78905	50099	42548	15209	14354	14238
(5) ④/①	1.2	10.2	9.4	15.6	10.6	15.5	6.0	7.4	10.2
(6) 処理時間	2294025.5	74412.4	127835.8	152377.3	40255.7	16762.0	15196.6	15613.5	11057.5
(7) DISK上の交換時間	8667.4	3223.9	18445.7	448.5	1642.7	267.2	1694.6	5817.0	2695.6
(8) ⑧-⑦	228538.1	7128.5	109390.1	151928.8	38613.0	16494.8	13502.0	9796.5	8358.9
(9) 転送処理時間	821395.2	27924.4	57753.3	106949.3	19821.8	7068.0	5675.6	3441.1	3615.2
(10) 照合・操作処理時間	0.8×107609 (35.9)	0.9×32173 (92.0)	1.9×31207 (52.8)	3.7×29027 (70.4)	4.6×12350 (51.3)	0.8×9141 (42.8)	0.8×6776 (42.0)	1.3×2650 (35.1)	0.9×3910 (43.2)
(11) プログラムの探索時間	40867.5	8625.3	11887.0	7370.3	1581.4	738.1	618.2	437.3	221.0
(12) プログラムの解放時間 (消去時間)	1.3×146175 (8.2)	0.7×11723 (12.1)	1.5×7777 (10.9)	1.5×4947 (4.9)	0.6×2362 (4.1)	0.4×1706 (4.5)	0.4×1440 (4.6)	0.7×635 (4.5)	0.3×664 (2.6)
(13) 数値演算時間	18302.4	337.2	6149.9	250.3	775.2	476.8	184.7	33.1	360.2
(14) 解放時間等 その処理時間	54630.2	3002.9	7144.8	21584.7	4488.1	1959.5	1237.8	712.8	832.2
(15) ⑮/①	0.4×142831 (2.4)	0.8×3663 (4.2)	1.0×7484 (6.5)	4.5×4839 (14.2)	1.2×3708 (11.6)	0.7×2048 (11.9)	0.8×1478 (9.2)	1.1×657 (7.3)	0.9×960 (10.0)
(16) 数値演算時間	29242.7	0	0	0	0	0	34.9	0	25.8
(17) 解放時間等 その処理時間	1.2×25004 (12.8)	29298.7	26455.1	15774.2	11746.5	6252.4	5750.8	5142.2	3304.5
(18) ⑮/①	2.1	2.0	2.0	3.1	2.5	2.3	2.3	2.7	2.4

COMPUTER" 大阪大学博士論文 (1976)

- (4) 緒方, 豊田, 田中: "SNOBOL インタプリタの評価" 昭和50年度電気関係学会関西支部連合大会予稿集 (1975)
- (5) 緒方, 三上, 豊田, 田中: "SNOBOL 処理システムの評価" 昭和51年度電子通信学会総合全国大会予稿集 (1976)
- (6) 緒方: "Disk および Cassette SNOBOL の動作評価" 昭和50年度研究会報告書 情報処理学会 関西支部 (1975)
- (7) 宮下: "LSP(FORTRAN)について" 昭和50年度研究会報告書 情報処理学会 関西支部 (1975)