

実行回数計数機能を追加した SNOBOL 4 処理系

吉田 和幸・牛島 和夫

(九州大学 工学部)

1. はじめに

ソフトウェア開発の初期の段階においてプロトタイプを組み立てて実際に動かしてみても問題点を認識する作業は有用である。解くべき問題領域が文字データ処理を中心とする場合などに SNOBOL 4 を使うと容易にプロトタイプの構築ができる [1, 2]。一方、プログラム中の各文の実行回数を知ることはプログラムの開発や改善に役立つ。SNOBOL 4 によるプロトタイプ作成を助けるために、既存の SNOBOL 4 処理系に実行回数計数機能を追加した。原著者の許可を得て入手したこの処理系は、SIL (SNOBOL 4 Implementation Language) という抽象言語で書かれたもの (OS 360 用、以下 SIL 版 SNOBOL 4 と呼ぶ) である [3]。これを FACOM OSIV/F 4 のもとに移し換える際に各文の実行回数を計数してソーステキストとならべて表示する機能を追加した [4]。さらにこの処理系を IBM VM/370 CMS と HITAC VOS 3 のもとに移し換えた [5, 6]。一方、漢字等を含む日本語ファイルの扱いを可能とする機能拡張も行った [7]。

本稿では、2 章で SIL 版 SNOBOL 4 とそれを記述している SIL システムの構成について簡単に述べる。3 章で実行回数計数機能の実現について述べ、機能追加作業が比較的容易に行えた理由を SIL 設計の観点から考察する。4 章では簡単なプロトタイプ作成の実例をとおして実行回数計数機能の評価を行う。5 章では実行回数追加版 SNOBOL 4 の移し換えについて触れる。日本語 SNOBOL 4 の詳細については別稿に譲る。

2. SIL 版 SNOBOL 4 処理系

SNOBOL 4 処理系はトランスレータ、インタプリタ、記憶管理、システムインタフェースという四つの部分から構成される [3]。トランスレータは SNOBOL 4 プログラムをプレフィックスコードに変換し、それをインタプリタが解釈実行する。記憶管理とシステムインタフェースはトランスレータとインタプリタの両方から使われる。記憶管理は領域の割り当て、データの記憶、記憶域の再構成等を行う。

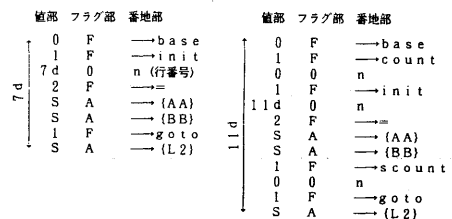
システムインタフェースは入出力やプログラム割込み等を受け持つ。

トランスレータが生成するプレフィックスコードはディスクリプタという単位からなっており、それは値部、フラグ部、番地部を持っている。例えば図 1 (a) の文は図 1 (b) のプレフィックスコードに変換される。フラグ部が F のディスクリプタは関数を表わし、その値部が引数の個数を、番地部が関数の入り口番地をそれぞれ表わす。フラグ部が A のディスクリプタはこの例では変数名 (ラベル名) をさすポイントを番地部に持つ (文字列等の実際の値を指すこともある)。init の引数のディスクリプタは番地部と値部だけが意味を持つ。番地部には文の番号が入り、値部にはその文の実行が失敗したときに読み飛ばすディスクリプタの数 (失敗オフセット) が入っている。

SIL 版 SNOBOL 4 処理系を記述している SIL という抽象言語は図 2 のような階層構成により実現されている [3]。最も外側 (図 2 (a)) が SIL の世界で 136 個の命令を持つ。その内側 (図 2 (b)) が 136 個の命令をアセンブラのマクロ機能を用いて定義した SIL マクロの層である。その内側に入出力等 OS に依頼する部分を IO ルーチンとしてまとめた層がある (図 2 (c))。これらのルーチンはアセンブリ語で記述され、SIL マクロから呼び出される。実際の入出力は FORTRAN の実行時ルーチンに処理を依頼し、異常終了時処理、時間管理等の OS に関係する部分は直接 OS のマクロ命令を用いているものもある。これらが最も内側の層を構成する (図 2 (d))。

L1 AA = BB :S(L2)

(a) SNOBOL 4 の実行文



(b) プレフィックスコード

(c) カウンタを導入したプレフィックスコード

図 1. SNOBOL 4 プログラムのプレフィックスコードへの変換例

3. 実行回数計数機能の実現

各文の実行回数を計測するにはソースプログラムを解析し、そこに

```
count (k) := count (k) + 1
```

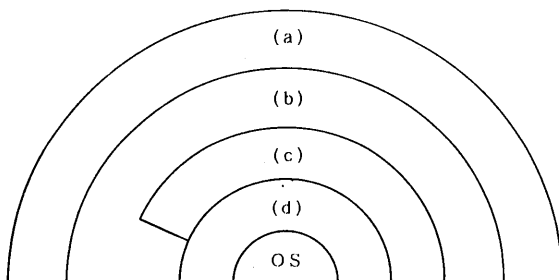
という形式のカウンタを挿入する方式（前処理／後処理方式）をとることが多い。FORDAP, COBOLDAP, PASDAP等はいずれもこの方式をとっている[8]。しかしSIL版SNOBOL4処理系は解釈実行方式をとっているため、この方法を採用すると挿入されたカウンタも解釈実行され計数のためのCPU時間、記憶場所の増加が著しい。このためSNOBOL4の実行回数計数機能はカウンタをソースコードではなくプレフィックスコードの形で挿入する方式をとることにした。SNOBOL4では各文の実行の成否によってプログラムの流れを変えることができる。従って、各文ごとに成功、失敗の数を計測することも重要である。そこで各文の実行回数とともにそれらも計測する。

実行回数計数機能を実現するためにSIL版SNOBOL4に次の追加修正を加えた。

- (1) カウンタの挿入 トランスレータがSNOBOL4ソースプログラムを変換するときに、文の実行の直前（init関数呼び出し前）に実行回数を計数する関数countを呼び出すプレフィックスコードを、文の実行部と分岐部の間に成功回数を計数する関数scountを呼び出すプレフィックスコードをそれぞれ挿入するようにした（図1.（c）参照。このためのトランスレータの書き替えはSIL言語で22行）。
- (2) ソーステキストの処理 実行回数をソーステキストと並べて表示するために、ソーステキストを一時ファイルに蓄える。トラン

スレータ中のソースリストを出力する部分を利用し、出力先を一時ファイルにすればよい。SNOBOL4ではセミコロンで区切って1行に複数の文を書くことを許している。そのような場合はセミコロンで分割して1行1文として保存するようにした。

- (3) カウンタ領域の確保 プレフィックスコードへの変換後、解釈実行に入る前にカウンタ領域を確保するようにした。このためSIL言語にカウンタ領域を確保する命令を追加し、これに対応して領域の確保をOSに実際に依頼するサブルーチンをIOLルーチンに加えた。
- (4) count, scountの追加 トランスレータが挿入したプレフィックスコードを実行するためにcount, scountの2つの関数をインタプリタに追加した（SIL言語で25行）。
- (5) 結果の出力 解釈実行後、ソーステキストとそれに対応する実行回数、成功回数、失敗回数を出力する。ソーステキストの分割等の処理は（2）で処理済みであるので、ここでの処理はソーステキストを読み、ソーステキストとそれに対応する実行回数、成功回数、失敗回数を並べて出力することである（SIL言語で61行）。
- (6) 実行回数計数機能使用の有無 計数機能を処理系に組み込んだのでTSSコマンド（あるいはジョブ制御文）のパラメータで使用の有無を選択できるようにした。このために実行時パラメータを判定する部分を書き直した。
- (7) 異常終了時の処理 実行回数計数中にプログラムが異常終了した場合、それまでに集めた実行回数情報はプログラムのデバッグ等に有用である。SIL版SNOBOL4ではオーバーフロー、記憶保護例外等のプログラム割り込みが起こった場合に、それまでの実行のサマリー（CPU時間、入出力行数等）を出力できる。実行回数計数版SNOBOL4ではプログラム割り込み時に実行回数情報を出力するほか、CPU時間切れ、TSS使用時の端末割り込み時にも実行回数情報を出力できるようにした。これを実現するためにトランスレータの実行に先立って1回呼び出されるSILマクロINITと終了処理を行うIOLルーチンFINISとの一部を書き替えた。すなわち、異常終了後直ちに後処理部に制御を移す指示をOSに登録することをINITで行わせ、その指示の解除をFINISが行う。CPU時間切れ時の処理に関し



- (a) SIL言語
- (b) SILマクロ定義
- (c) IOLルーチン
- (d) OSのマクロ命令, FORTRAN実行時ルーチン

図2. SIL処理系の階層構成

てはスーパーバイザマクロを用いて直接OSに依頼している。端末
割り込み時の処理は、FORTRANの実行時ルーチンを利用し
ている。

これらの追加修正をまとめる。

・(1), (2), (4), (5)はSIL言語(図2(a))で書かれた部分(全体
で約6800行)の修正である。修正量は合計で約280行である。

・(3)の前半はSILマクロ定義(図2(b))の追加である。この部
分は6行である。

・(3)の後半と(6), (7)はIOLルーチン(図2(c))での修正である。
IOLルーチンはモジュールの集まりであり、それらのモジュール間で
の副作用はない。ここでの修正はモジュールの置き換え、追加であり、
IOLルーチン全体で2260行のうち、修正量は160行である。

SIL版SNOBOL4は約10年前に設計実現されたにもかかわらず、
実行回数計数機能の追加作業をこのように比較的容易に遂行で
きたのは、ドキュメントがよく整備されていたこと、SILシステム
の設計が明快でよく機能分割されていたことが大きな理由と考えられ
る[3]。

(1) SILシステムの構成 第2章にまとめたように構成が透明で
機械またはOSに独立な部分と依存する部分が明確に分離されて
いる。従って一部の修正がプログラムの広域に影響を及ぼすこと
がなかった。

(2) データの表現 SILプログラム中のデータはディスクリプタ
単位で数える。このためSILマクロ定義の段階でバイトマシン、
ワードマシンの違い等を吸収できる。

(3) プレフィックスコード プレフィックスコードの構造やSNO
BOL4ソーステキストからの変換の規則がドキュメントから容
易にわかるので、カウンタの挿入場所の決定が容易にできた。

(4) SIL言語の定義 SIL言語の文法に関するドキュメントに
よりSIL言語の修得が容易であった。そのためカウンタの挿入、
結果の出力等の追加部分のほとんどをSIL言語で書けた。さら
にSIL言語の拡張の際にも既存のSILの機能と矛盾すること
なく拡張できた。

(5) IOLルーチンの仕様 外部仕様がドキュメントによりはつきり
わかっているので書き替えが必要なルーチンは容易に書き替えら
れた。さらにサブルーチン名、大域変数名等のSILで書かれた
SNOBOL4本体と連絡する名前が明示されているので、作業
変数等の追加変更が自由にできた。

4. 実行回数計数機能の評価

実行回数計数機能の使用例を図3に示す。図の右に各実行文の実行
回数、成功回数、失敗回数が出力されている。このプログラムはFO
RTRANのソーステキストを入力し、それを構成するプログラム単

SOURCE STATEMENT		EXECUTION	SUCCESS	FAILURE
* LIST THE LINE NUMBER OF EACH FORTRAN PROGRAM UNIT		00000010		
* AUTHOR. DR. N. FUJIMURA.		00000020		
* DATE-WRITTEN. 80/10/03.		00000030		
* LINE NO = 0		00000040		
LINE = '***MAIN**'		00000050		
STARTLINE = 1		00000060	1	1
OUTPUT = ' FROM TO PROGRAM NAME'		00000070	1	1
OUTPUT =		00000080	1	1
* INPUT BUF = INPUT :F(END)		00000090	1	1
LINE NO = LINE NO + 1		00000100	1	1
BUF LEN(72) . BUF		00000110	1	1
BUF (POS(0) 'C' ABORT) I		00000120	1	1
BUF (POS(5) (' ' I 'O'))	:S(ST)F(INPUT)	00000130	1640	(a) 1639
ST BUF 'FORMAT' :S(INPUT)		00000140	1639	1639
BUF 'END' :S(OUTPUT)		00000150	1639	1639
BUF (' SUBROUTINE' I 'FUNCTION' I		00000160	1639	(b) 987
' BLOCK ')	:S(PROG)F(INPUT)	00000170	987	28
* PROG LINE = BUF		00000180	959	(c) 30
DLTBLK LINE FENCE ' ' =	:S(DLTBLK)F(INPUT)	00000190	929	(d) 29
* OUTPUT OUTPUT = ' ' DUPL(' ',6) - SIZE(STARTLINE)) STARTLINE		00000200	29	29
' ' DUPL(' ',6) - SIZE(LINE NO)) LINE NO		00000210	29	174
' ' LINE		00000220	29	29
STARTLINE = LINE NO + 1		00000230	30	30
LINE = '***MAIN**'	:S(INPUT)	00000240	30	30
* END		00000250	30	30
		00000260	30	30
		00000270	30	30
		00000280	30	30
		00000290	30	30
		00000300	30	30
		00000310	30	30
		00000320	1	1

図3. 実行回数計数機能使用例

位の名前とその区切りを表示するプログラムである [9]、図3

(a) から入力となったFORTRANプログラムの総行数が1639行であったこと、(b) からコメント行と継続行を除いた行が987行であったこと、(c) からEND行が30行であったこと、(d) からサブルーチン文等が合計29あったことがわかる。

このプログラムの実行には実行回数を計数して約5.5秒かかる。1600行余のプログラムを処理するための5.5秒は長すぎるので実行回数情報を手掛かりに実行時間の短縮を試みる。(b)、(d)に対応する実行文はそれぞれ一行で多くの仕事をしようとして複雑なパターンを使っている。しかもこれらの文は(b)が1639回実行され、(d)も929回実行されている。そこでそれぞれ簡単なパターンを使った複数の文に置き換えて以前と同じ入力データを与えたものを実行回数情報つきで図4に示す。図4のプログラムの実行は約2.8秒で済み、実行時間はほぼ半分に短縮された。

FORTRANの文は72ヶすべてを使うことはまれで、1行の右の方は空白であることが多い。そこで図4のプログラムの入力部でSNOBOL4のTRIM機能を用いて1行の右側の空白を取り除き、パターンマッチの際の探索範囲を狭めたところ、実行時間は約1.8秒となり、図4のプログラムに対して約60%実行時間が短縮した。表1にこの3個のプログラムの実行時間を示す。表1では実行回数を計測した場合とともに計測をしない場合の実行時間、およびそれらの比を合わせて示している。実行回数を計測した場合の実行時間の増加は1割以内に収まっている。

図3のプログラムは「FORTRANソーステキスト中からサブルーチン等を切り出したい」という要求[9]があった時に30分ほどで作成したプログラムである。SNOBOL4は挿入、削除を伴ったパターンマッチを一文で書け、その組み合わせにより文字列に関するかなり複雑な操作も簡単に記述できる。従ってテキストファイル処理

表1. 各プログラム実行時間

プログラム	実行回数計数 (P) [ミリ秒]	計数なし (NP) [ミリ秒]	比 (P/NP)
図3	5552	5430	1.022
図4	2866	2753	1.041
図5	1771	1665	1.070

などに際して適当なツールがない場合その場でプログラムを作ってみるというラビッドプロトタイピングに適した言語である[10]。このような機能のツールを頻繁に使用する場合には他の言語を用いて高速版を作り直せばよい。しかし1.8秒ならばSNOBOL4版でも十分実用になる。この場合SNOBOL4の実行回数計数情報は、入力データに関するドキュメントとして利用でき、作成したばかりのSNOBOL4プログラムの正しさを確かめる手掛かりになる。

SNOBOL4における実行回数計数機能の使用上の効用をまとめる。

- (1) 作成したプログラムの動作の客観的認識が可能である。
- (2) 採用したアルゴリズムの適否を実行回数から判定可能である。
- (3) 実行箇所と非実行箇所が一目瞭然なので、プログラムの検査に役立つ。
- (4) 実行回数がプログラムの入出力テキストに関するドキュメントとして利用できる。
- (5) 実行回数を手掛かりにプログラムの効率改善に役立つ。
- (6) カウンタをプレフィックスコードの形で挿入するため実行回数計数に要する時間の増加はわずかである。

5. 実行回数計数機能追加版の移し換え

SIL版SNOBOL4は現在40以上の機種で稼働している。実行回数計数機能追加版もできるだけ多くの機種の上で動くことが望ま

ST	BUF FENCE 'C'	:S(INPUT)	00000160	1639	586	1053
	BUF FENCE LEN(5) ' '	:S(ST)	00000170	1053	987	66
	BUF FENCE LEN(5) '0'	:S(ST)F(INPUT)	00000180	66	0	66
	BUF ' FORMAT'	:S(INPUT)	00000190	987	28	959
	BUF ' END '	:S(OUTPUT)	00000200	959	30	929
	BUF ' SUBROUTINE'	:S(PROG)	00000210	929	24	905
	BUF 'FUNCTION'	:S(PROG)	00000220	905	4	901
	BUF ' BLOCK '	:S(PROG)F(INPUT)	00000230	901	1	900

図4. パターンマッチ部分の改良版 (部分)

	&TRIM = 1		00000120	1	1	0
	INPUT('SYSPIT',5,72)		00000130			
			00000140	1	1	0
			00000150			
INPUT	BUF = SYSPIT	:F(END)	00000160	1640	1639	1

図5. TRIM機能使用による改良版 (部分)

しい。手初めに互換機といわれるIBM機、HITAC Mシリーズ機への移し換えを試みた。

5. 1. OS360からOSIV/F4への移し換え

我々が入手したSIL版SNOBOL4はOS360用のものであったので実行回数計数機能を追加するに先立ってまず九大型計算機センターのFACOM OSIV/F4上に移し換える必要があった。

SNOBOL4処理系が入出力に用いているFORTRAN実行時ルーチンの呼び出し方法がOS360とOSIV/F4とは異なっているため、入出力に関係するSILマクロとIOルーチンとを変更せねばならない。

SILマクロでは、FORTRANの実行環境を整えるINIT、出力の一部を行うOUTPUTとそれに関する命令、リワインド等の補助入出力命令を実行するREWIND等、計10個のマクロを変更した。

IOルーチンでは、入力を行うSTREAD、大部分の出力を行うSTPRNT、終了処理を行うFINISの3ルーチンを変更した。

OSIV/F4への移し換えは、SNOBOL4処理系本体を詳しく理解することなしに、ドキュメントが整備されているSILマクロとIOルーチンの変更のみで済んだ。

しかし、FORTRAN実行時ルーチンの呼び出し方法がOS360とOSIV/F4とで違うという事実気付くこと、およびOSIV/F4のFORTRAN実行時ルーチンの呼び出し方法を知ること到大変苦労した。OS360とOSIV/F4は互換性のあるそれぞれのハードウェア上で動作し、OS自体も互換性があるとされているため、FORTRAN実行時ルーチンも互換性があるものと単純に考えていたためである。しかし、互換機といってもOSも違い、FORTRANコンパイラ、実行時ルーチンも違う。一般にユーザが直接扱うことのない実行時ルーチンの仕様が異なるのは当然ありうることであった。OSIV/F4のFORTRAN実行時ルーチンの使用方法に関する情報は一部がマニュアル中に散見される程度で整理して提供されていない。そのためFORTRANコンパイラの出力をダンプする等試行錯誤したが、よくわからず、結局富士通の当事者に訪ねたところすぐに解決し、上に述べた様にわずかな修正で動作した。

5. 2. VM/370 CMSとVOS3への移し換え

OSIV/F4上で実行回数計数機能を追加したSNOBOL4処理系をIBM VM/370 CMSとHITAC VOS3とに移し

換えることは比較的簡単であった。アセンブリ言語がOSIV/F4とCMS、VOS3とで同等であり、FORTRAN実行時ルーチンはCMSとOS360がほぼ同じ、CMSとVOS3が同じであったからである。OSIV/F4への移し換える経験からFORTRAN実行時ルーチンの呼び出し方法にCMSとOS360、OSIV/F4とで違いがあることが予想されたので、初めからそのつもりでその調査にあたったためでもある。

異常終了時の処理は、オーバーフロー、記憶保護例外等に対してはOSIV/F4とCMS、VOS3とで同じ仕様のスーパーバイザマクロを用いるので、それが起こった場合の実行回数計測結果の出力は可能である。CMSは会話型の単一ユーザ用OSであるため、CPU時間切れ等による異常終了は原則として起こらない。しかし、OSIV/F4で実現されている端末割込み時の計測結果の出力はCMSとVOS3とでは実現できなかった。これは、OSIV/F4 FORTRANの実行時ルーチンではその初期設定の際に端末割込み時の処理のためのルーチンを設定できるが、CMSとVOS3とではこのことが実現できるかどうか調査の方法がなかった。OSIV/F4でもこのことはマニュアルには書いてなく、実行時ルーチンをダンプして初めてわかったものである。CMSやVOS3は手元の計算機ではないのでダンプのような方法はとりにくかった。

この移し換えでは実行回数計数機能を実現するためにOSIV/F4上でSNOBOL4処理系に追加した部分は全く変更する必要がなかった。

なお、VOS3への移し換えは1981年末から利用可能になったN1方式大学間ネットワークを利用して九大型計算機センターから東大大型計算機センターへ行ったものである。ネットワークを使用したための移し換え上の問題点については[6]を参照されたい。

5. 3. 移し換えにおける問題点

実行回数計数機能追加版SNOBOL4を実現したFACOM M200とそれを移し換えたHITAC M200H/M280H、IBM 370/138とは互換機であるといわれ、アーキテクチャは基本的に同じである。しかし、OSが異なり、FORTRANコンパイラ、実行時ルーチンが異なるため、今回のような問題が生じた。

SNOBOL4処理系が使用するOS等の支援ルーチンは次の2つに分けられる。

(1) OSの機能を直接用いるもの 異常終了処理等はスーパーバイ

ザマクロを用いて直接OSに要求を出す。これらのマクロは3機種間でほとんど違いがなく、ほぼそのまま動作した。しかもスーパーバイザマクロ等に関しては3機種ともマニュアルがしっかりしているのです。それらの比較も容易である。

- (2) FORTRAN実行時ルーチンを用いるもの SNOBOL 4の入出力はすべてFORTRAN実行時ルーチンを通して行われる。FORTRAN実行時ルーチンが用いているOSのデータ管理マクロは3機種ともほとんど同じであるにもかかわらず、当該ルーチンには互換性がない。しかも、このルーチンはFORTRAN利用者が明示的に用いるものではないので、それに関するドキュメントが（存在したとしても）ほとんど入手できない。そのため、SNOBOL 4処理系の移し換えに当たってFORTRAN実行時ルーチンの仕様が問題となった。これがはっきりわかってしまえば、5. 1. で述べたように移し換えに当たっての変更箇所はわずかである。

6. むすび

SIL版SNOBOL 4処理系に実行回数計数機能を追加し、さらに2つの計算機システム上に移し換えた。SILの設計は元来、多くの計算機システムの上でSNOBOL 4処理系を実働させることが第一義であったものと思われる。しかしその設計の故に、機能拡張を容易に行うことができたことを強調しておく。

なお、現在OSIV/F 4の上で動く実行回数計数機能版SNOBOL 4は九州大学大型計算機センターのほか、名古屋大学大型計算機センター、大阪大学社会科学研究所、慶応大学情報科学研究所でも稼働している。

参考文献

- [1] Zerkowitz, M.V.: A Case Study in Rapid Prototyping, Softw. Pract. exp., Vol.10, pp.1037-1042, 1980.
- [2] Fleck, A.C.: Verifying Abstract Data Types with SNOBOL4, Softw. Pract. exp., Vol.12, pp.627-640, 1982.
- [3] Griswold, R.E.: The Macro Implementation of SNOBOL4, W.E.Freeman & Co., p.310, 1972.
- [4] 牛島, 高比良: SNOBOL 4輪郭作成機能の使用について, 九大大型計算機センター広報, Vol. 12, No. 2, 1979.
- [5] 牛島, 吉田, 高比良: 実行回数計数機能を追加したSNOBOL 4処理系とその移植および使用について, 第21回情報処理学会全国大会論文集, pp. 223-224, 1980.
- [6] 吉田, 牛島: N1ネットワークを経由したソフトウェアの異機種間移し換え, ソフトウェア工学研究会資料25-4, 1982.
- [7] 牛島, 黒坂: 日本語文字処理機能を追加したSNOBOL 4処理系, 第25回情報処理学会全国大会論文集, pp. 255-256, 1982.
- [8] 牛島, 藤村: 実行回数計数ツール実現の総括, 第21回プログラミング シンポジウム報告集, 1980.
- [9] 牛島, 田町: プログラム移換作業から見たプログラミング環境の評価について, ソフトウェア工学研究会資料17-12, 1981.
- [10] 木村: SOFTWARE TOOLとは何か, 情報処理, Vol. 20, No. 8, pp. 673-680, 1979.