

述語の動作例を用いてリスト処理プログラムを合成する手法について

仲瀬明彦 門倉敏夫
(早稲田大学理工学部電気工学科)

深沢良彰
(相模工業大学工学部情報工学科)

Progプログラム合成の一手法として、プログラムの具体的な動作を記述した述語の例からProgのリスト処理プログラムを合成する手法と、それを実現するシステムについて述べる。

1. 本システムの概要

帰納的推論[1]とは、幾つかの具体的な例からより一般的な法則を推論することをいう。プログラム合成の分野では、プログラムの動作を記述した具体的な入出力の例を与えて、その例の示している動作をより一般的に行なうプログラムを合成する作業が帰納的推論の一例と考えられる。

プログラムの具体的な入出力例からプログラムを合成する手法は、リスト処理の分野では、対象言語をLispに設定して数多くの研究がなされてきた[2]-[4]。また近年、知識情報処理技術の研究の発達等により、Lispで作成できるプログラムもProgを用いて作成する要求、利点も論議されてきた[5]-[7]。このような背景により、Progプログラムを述語の具体的な動作例から合成するシステムもいくつか研究されている[6], [8]。

Progのプログラムを具体的な動作例から合成する研究には大別して2つのアプローチがある。1つは、述語内の引数を入力リストと出力リストのペアとして考えた、複数の述語の動作例を与え、それらの入力リスト間の再帰性と出力リスト間の再帰性を発見し、プログラムを合成するもの[8]である。もう1つは、与えられた述語の動作例から、先ずモデルとなる大まかなプログラムを推論し、次にユーザーとそのプログラムに対しての対話をを行いながらプログラムを修正し、ユーザーの望むプログラムを推論してゆくもの[6]である。

以上で述べた研究では、単一の関数や述語を合成することに関してはかなり良い結果が示されている。

以上に示したような帰納的推論アルゴリズムを用いてプログラムを合成する場合、1回の合成で作成できるプログラムはごく小さなもの（小さなサブルーチン、関数、述語等）で、大きなプログラムを合成することは出来ない。そこで、それらの手法を用いて大きなプログラムを作成する場合は、プログラムの下位部品を作成する支援システムとして使用されることが効果的であると思われる。しかし、実際に使われているプログラムの下位部品の作成支援を行うためには、帰納的推論で合成できるプログラムの対象範囲を広くしなければならず、そのことによる合成に必要な時間の増大や、合成に必要な動作例の増大は避けられない。

本システムでは、单一の述語の動作例の中で引数として与えられているデータのリストを変形することにより、その変形過程の再帰性を発見し、プログラムを合成する手法をとっている。この手法をとることにより、リストの中のアトムの意味の違いや、入出力、数値データに対する扱いも簡単になるので、幅広い分野のプログラムが合成できる。また、データの変形や再帰性発見においては、完全な探索やマッチングは行わず、統計的に実際のProgプログラム中で頻繁に使われているパターンのみについての探索やマッチングを行う。これにより、合成の完全性は無くなるが、適用分野の広さに比べて合成に必要な時間(計算量)や、合成に必要な述語の動作例が少なくて済む。

以下に、本システムの用いているプログラム合成のアルゴリズム、実際のプログラム合成の例、合成されたプログラムを用いたより大きなプログラムの作成支援の例について述べる。

2. 再帰性とプログラム合成について

再帰的なプログラムは帰納的推論を用いて合成しやすいので、多くのシステムがプログラムの再帰性を発見して再帰的プログラムを合成する手法をとっている[2], [8]。本システムでもこの手法を用いることにし、いかにして再帰的なプログラムを合成するかを以下に述べる。

再帰的な Prolog プログラムの実行中においては、入力した定数データの動作も又、再帰的な動きをしている。

(例) 2つのリストを結合するプログラム。

```
append([], X, X).  
append([A;X], Y, [A;Z]) :- append(X, Y, Z).
```

において、質問

```
?-append([a, b], [c, d, e], [a, b, c, d, e]).
```

を入力し、プログラムの振舞いをトレースすると、以下のようになる。

```
append([a;[b]], [c, d, e], [a;[b, c, d, e]])  
:-append([b], [c, d, e], [b, c, d, e]). (1)  
append([b;[]], [c, d, e], [b;[c, d, e]])  
:-append([], [c, d, e], [c, d, e]). (2)  
append([], [c, d, e], [c, d, e]). (3)
```

ここでデータのみに注目すると、以下のようになる。

| | 第一引数 | 第二引数 | 第三引数 |
|-----|---------|-----------|------------------|
| (1) | [a;[b]] | [c, d, e] | [a;[b, c, d, e]] |
| (1) | [b] | [c, d, e] | [b, c, d, e] |
| (2) | [b;[]] | [c, d, e] | [b;[c, d, e]] |
| (2) | [] | [c, d, e] | [c, d, e] |
| (3) | [] | [c, d, e] | [c, d, e] |

上の例からも明らかなように、データのみに注目しただけでも再帰的な規則性が発見できる。

上の例では、最初からプログラムが存在しており、これにデータを入力し実際にプログラムを実行させることにより、データの規則的な変形の履歴を得ている。ここで、この操作を逆に辿ってプログラムを得ることを考える。つまり、データとデータの規則的な変形の履歴を最初に与えておき、それよりプログラムを合成するのである。

従って、本システムがサポートしているデータ変形規則を用いても定数データが変形できなかったり、データ変形の履歴から再帰性が発見できなかつた場合は、プログラムの合成は失敗する。

3. システム構成

図 1 にシステムの構成を示し、各部の詳細を以下に示す。

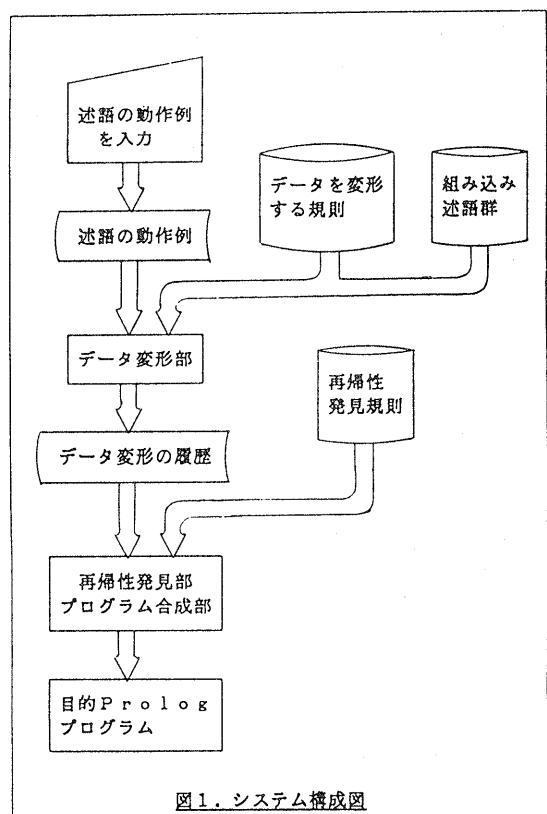


図 1. システム構成図

(1) 述語の動作例の入力

合成したい述語の具体的な動作を記述したもの を入力する。述語内の引数（データ）は、アトム、リスト又は数値でなければならない。但し、引数の数に制限はない。

例えば、リストの反転を取るプログラムを合成したい場合は、

```
reverse([a, b, c, d, e, f, g], [g, f, e, d, c, b, a]).
```

などという動作例の入力が考えられる。

また、以下に本システムにおいて拡張された、述語の動作例の表現方法を示す。

(a) リスト内の要素の意味の違いを扱う場合。

入力文字列のなかである記号が発見された時にその次の文字を削除するプログラムを合成したい場合次の例が考えられる。

```
del([a, b, c, d, e, f, *, g, h, i],
```

```
[a, b, c, d, e, f, *, h, i]).
```

ここでは、"*" のあとの文字を削除する述語を具体的な例で示している。この例では、リスト中の "a, b, c, …" はリストの構造を記述するためのアトムとして使われているが、"*" は前者の意味に加えて次のアトムを削除する特殊文字としての意味も持たされている。このようにリスト内のいくつかの要素に特殊な意味を持たせたい場合は、その要素を "{}" で囲んだ動作例を入力する。つまりこの場合は、

```
del([a, b, c, d, e, f, {*} , g, h, i],
```

```
[a, b, c, d, e, f, {*} , h, i]).
```

という動作例を入力する。

1 つのリスト中でこのような指定を複数回行ないたい場合は {} の後に添え字をつけて識別する。

例えば、* の次の二文字と # の前の二文字を削除するプログラムを合成したい場合、

```
del2([a, b, c, d, {*}1, e, f, g, {#}2, h, i, j, k, l],
```

```
[a, b, c, d, {*}1, f, {#}2, h, i, j, k, l]).
```

と入力する。

(b) プログラムの実行中に入出力を行ないたい場合

入出力は、シーケンシャルに行われるを考え、目的のプログラムが呼出されてからの入出力にもシーケンシャルな通し番号を付け、その番号を n とする。

入力は、inp(n) (入力ストリームからの入力データ)、出力は、outp(n) (出力ストリームへの出力データ)、で示される。入出力はこの通し番号

の順番で行なわれる。入出力ストリームをプログラムの実行中に変化させる事は、現在はサポートしておらず、それらはキーボードからの入力とディスプレイへの出力に設定されている。

例えばリストが与えられており、次に数字 n を読み込み、リスト内の要素でリストの最初から n 番目にある要素を書き出すプログラムに対しては、

```
writeatom([a, b, c, d, e, f, g, h, i]),
```

```
inp(1)(3), outp(2)(c).
```

という動作例を入力する。

(2) データを变形する規則とその適用方法

本システムでは、付録 1 のデータ変形規則に従って述語の動作例内の引数データを変形する。付録 1 のデータ変形規則は必ずしも必要十分なものではないが、Prolog のプログラムの例が多く記載されている文献中 [9], [10] で頻出している Prolog の事実 (Fact) や規則 (Rule) を優先的に抽出し、それと等価に働くデータ変形規則に置き換えたものである。以下の図 2 は、リスト処理の Prolog プログラム中で使われている最下位の述語 70 個をサンプリングした結果を示している。

| | | |
|-----|--------------------------------|----|
| (1) | 述語内の引数が全て空リストになると終了 | 7 |
| (2) | 述語内の引数の中で、空リストでない変数が 1 つになると終了 | 4 |
| (3) | 述語内の引数の中で、リストの頭部又は尾部が等しければ取り除く | 13 |
| (4) | 述語内のいくつかのリストを、頭部と尾部に分ける | 9 |
| (5) | (1)-(4) の操作を行い、且つ入出力を行う | 1 |
| (6) | (1)-(4) の操作を行い、且つ数値を 1 減らす | 2 |
| (7) | 数値が 0 になると終了 | 2 |
| (8) | 本システムの組み込み述語を適用して終了 | 11 |
| (9) | その他 | 21 |

図 2. Prolog プログラム中で発生するパターン

付録1のデータ変形規則を総あたりで述語の引数データに適用すれば、必ず一つは目的のデータ変形履歴が得られるが、なるべく高速にデータ変形履歴を得るために付録2に示すアルゴリズムでデータ変形規則を適用する。

また図3は、付録1のデータ変形規則中で使われている本システムの組み込み述語を示したものである。

| | |
|--------------------|------------------------|
| equal(X, X, ...). | 引数の値が同じ |
| car([X;Y], X). | 片方の引数の頭部が他の引数と同じ |
| cdr([X;Y], Y). | 片方の引数の尾部が他の引数と同じ |
| cons(X, Y, [X;Y]). | 1つの引数の頭部と尾部が他の2つの引数と同じ |

図3. 本システムが用いている組み込み述語

(3) データ変形履歴からのプログラムの合成

プログラムの合成は、(2)で作成されたデータ変形履歴を用いて以下のステップに基づいて行われる。(以下で現れる"述語記号K"とは、最初に述語の動作例として入力した述語の述語記号に数字Kを付けたものをいう。)

<1> データ変形履歴の述語形式への変換。

このステップは、以下の手順で行われる。

(1) "述語記号1

(<規則が適用出来なくなった時のデータ>)."-を登録する。

(2) データ変形履歴に最後に登録されたものから順番に、規則を適用した各段階ごとに、

"述語記号N+1

(<規則を適用する前のデータ>):-

述語記号N

(<規則を適用した後のデータ>)."を登録する。

(3) "最初に例として入力した述語:-

述語記号Nmax

(<例として入力した述語のデータ>)."を登録する。

数値を扱っている場合は数値を1減らす述語を、入出力を行う場合はread文やwrite文を、それぞれ":-"の後に挿入する。

<2> 定数記号の変数化。

<1> で出来たプログラムの各述語(又は述語から述語への規則)の中で、同値のリストや数値が使われている場合は、データを変形した規則を参照しながら、そのリストや数値を変数に置き換える。但し、3.(1).(a)に規定されている特殊な意味をもつ要素に対しては、変数化は行わず定数として扱う。

<3> 再帰性発見、プログラム合成

<3> のプログラムから再帰性を発見し、入力した述語の動作例の引数データ以外のデータに対しても働くプログラムを作成する。隣り合った節が同じパターンの規則を用いている場合、そこが再帰性をもっていると考える。それら複数個の隣り合った述語の述語記号を固定した述語記号に置き換え、隣接している述語の最上端と最下端のみの述語を残す。つまり、<2>で出来たプログラム中に節(P2:-P1, P3:-P2, ..., Pn+1:-Pn)が同じパターンの述語である時、述語記号の固定された述語をPとすると、P:-P1, P:-P, Pn+1:-Pとする。この手続きをそれ以上適用出来なくなるまで、繰り返し適用する。

4. システムの使用法

本システムは、ユーザーの入力した具体的な動作例を1つ読み込み、合成を行う。システムによって合成されたプログラムはユーザーが適当なテストを行うように設計されている。また、データ変形規則からも分かるように、規則(3), (4), (5), (6), (9)が非決定的な規則となっており、複数のデータ変形の履歴が生成される。もし最初のデータ変形履歴から合成されたプログラムがユーザーの意図しないものであれば、システムは次のデータ変形履歴から再度プログラムを作り直す。

(1) リスト中の最初から数えて n 番目にある要素を取り出す述語 : `substring(X, N, A)`.

入力した述語の動作例。

```
substring([a,b,c,d,e], 3, c).
```

(2) リスト中のあるアトム又はリストを他のアトム又はリストに置き換える述語 : `subst(X, A, B, Y)`.

入力した述語の動作例。

```
subst([a, b, c, d, e, f, g], c, x, [a, b, x, d, e, f, g]).
```

(3) リスト中に"is"というアトムがあった場合そのアトムとその1つ前のアトムを入れ替える述語 : `quest(X, Y)`.

入力した述語の動作例。

```
quest([he, {is}, a, nice, boy],  
      [{is}, he, a, nice, boy]).
```

図4. 本システムを用いて合成できた述語の例

5. システムの評価と応用分野

本システムを用いての小さなプログラムの合成の例を付録3に示す。本システムをプログラムの合成に使用した結果、かなり広い範囲の有用なリスト処理用プログラムが合成できることが分かった(図4)。また、合成を行った例に対しては、ほとんど1回の試行で目的プログラムを合成出来た。

小さなプログラムを合成することに対して有用であることは勿論であるが、大きなシステムを作成する場合の支援システムとしても利用できる。このシステムを用いて簡単な行エディタの作成をした例を付録4に示す。

6. 今後の課題

(1) 作成されたプログラムのテストは、人手で1つ1つテストデータを入れてテストしなければならないが、システムの方で典型的な述語の入

出力データを示してくれるようになる。これにより、テストが容易になる。

(2) 現在は単一の述語の動作例の中のリストの要素の再帰性にのみ注目しているが、複数の述語の動作例の間の関係も発見できるようにする。

(3) 述語の動作例の入力に於いて、真になる動作例だけでなく、偽になる動作例も入力できるようにする。(2)と(3)により、合成できるプログラムの適用分野が一層広がると思われる。

(4) 付録1のデータ変形規則と付録2の再帰性発見アルゴリズムにより多くの規則をもたせれば、合成できる述語の範囲は広がるが、どの程度までそれらを拡張すれば合成に必要な計算時間の増大に比べて効果的な拡張ができるか考察してゆく。

参考文献

- [1] Angluin, D and Smith, C. H.: Inductive Inference Theory and Methods, Computing Surveys Vol. 15, No. 3 (1983).
- [2] Summers, P. D.: A Methodology for LISP Program Construction from Examples, J. ACM Vol. 24, No. 1 (1977).
- [3] Hardy, S. : Synthesis of Lisp Functions from examples, IJCAI (1975).
- [4] Shaw, D. E., Swartout, W. R. and Green, C. C. : Inferring Lisp Programs from Examples, IJCAI (1975).
- [5] Computer Today (1984. 1).
- [6] Shapiro, E. Y.: Algorithmic Program Debugging, MIT Press (1982).
- [7] 情報処理 Vol25. No. 12 (1984).
- [8] Yokoi, S. : Inference of Programs from Examples. 情報処理学会29回全国大会
- [9] 安部： Prologプログラミング入門 共立出版 (1985).
- [10] Clocksin, C. S. and Mellish, C. S. 中村訳： Prologプログラミング (1983).

付録1. 述語内のデータを変形する規則の概要

- <再帰の修了条件となるもの>
- (1) 例内のデータが全てリストになつたら、規則の適用を終了する。
(`Prolog`では、`pred([],[],[])`に相当)
 - (2) 例内のデータに規則を適用できなくなったとき、規則の適用を終了する。条件としては、複数個ある最初に与えられた引数としてのデータの中で、空リストでないものが1つとなつた場合である。)
 - <基本リスト処理手続き>
 - (3) 例内のデータ中のn個のデータの頭部が同じならそれを取り除く。
(`Prolog`では、`pred([A|X],[A|Y],...,[Z]) :- pred(X,Y,...,Z).`に相当)
 - (4) 例内のデータ中のn個のデータの尾部が同じならそれを取り除く。
(`Prolog`では、`pred([A|X],[B|X],...,[K|X]) :- pred(A,B,...,K).`に相当)
 - (5) 例内のデータ中のあるデータを頭部と尾部に分ける。
(`Prolog`では、`pred([(A|B)|C,...,(K|L)|M|N|...|O|P|Q|...|R|S|T|...|U|V|W|...|X|Y|...|Z]) :- pred2(A,B,C,...,K), pred2([A|B],...,K).`に相当)
 - (6) 例内のデータ中の2つのデータを頭部と尾部に持つデータを作る。
(`Prolog`では、`pred(A,B,...,K) :- pred2([A|B],...,K).`に相当)
- 〔注〕(5)の規則は、繰り返し適用すると、述語の引数の数が増大する一方で、そこで、述語を発見することの妨げとなる。したがって、データの変形の過程では、データの個数はn+1個までしか増大しないような制約を設ける。(5)の規則を繰り返し適用する場合には、(5)と(6)を述語の例を入力し、規則の適用を行なう。そのため、以前に規則(5)により分離されたデータを尾部に持ち、規則(6)によって分離されるデータを頭部に持つデータムを作成す。
- `Prolog`のプログラムで表わすと、
`pred(A,[P|B],C,...,K) :- pred([P|A],B,C,...,K).`と等価である。]
- <入出力に関するもの>
- (7) 対象された述語の動作例内に`outp(n)(X)`がある時、例内のデータ中のあるデータに対し(3)、(4)を行い、かつその操作で分離されたものが`outp(n)(X)`であれば、Xを出力する。
 - (8) ある項を入力して例内のあるデータの頭部に結合する。
(`Prolog`では、`pred(X,[..],Z).`と
`d(A),pred([A|X],[..],Z).`に相当)
- 〔注〕数値を入出力する場合には、各述語内の引数の数を1つ増やすか減らさずとする。
- 各々 `pred(A,...,N)` :>
[read(X),pred(X,A,...,N).]
`pred(A,B,...,N)` :>
[read(X),pred1(X,A,...,N).]
`pred1(A,B,...,Y)` :>
[write(A),pred1(B,...,Y).]
に相当する。]
- <数値に関するもの>
- (9) 例内のデータ中に数値がある場合、(3)、(4), (5)、(6)を行い、数値データの値を1減らす。
(`Prolog`では、`pred(X,M,...,Z) :- N is M-1, pred(X,N,...,Z).`に相当)
 - (10) 例内のデータ中の数値データの値を含む述語に対しては、数値データの値を同時に1減らすことによって、数値データを減らしてゆき0にしようとする試み。
- (10)例内のデータ中の数値データの値が0であれば、その引数に対する操作を終了する。
(`Prolog`では、`pred(A,B,0,...) :- pred(A,B,0,...).`に相当)
- <組み込み述語に関するもの>
- (11)動作例`pred(X1,X2,...,Xn)`に対してTの値を持つ述語を動作例に適用し、規則の適用を行なう。
- 組み込み述語に用意している組み込み述語は、`LISP`の最も基本的な関数を使用しており、付録2に示してある。】

付録2. データ変形規則を適用するアルゴリズム

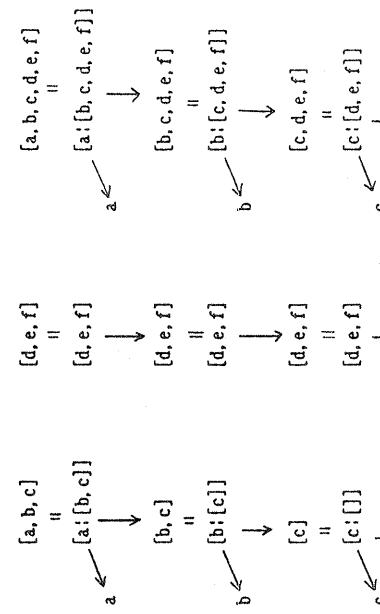
```

procedure history(データ並び);
  if (1) or (2) が成立 then
    begin 履歴ファイルに記録;
      return;
    end
  else if (11) が成立 then
    begin 履歴ファイルに記録;
      return;
    end
  else if データ並びの中にinp(n)(A)又はoutp(n)(A)に対しても小さい inp又は output は存在しない(7)又は(8)を行なう;
    begin n の最も小さい inp 又は output に對して(7)又は(8)に記録;
      履歴データ並び:=データ並びを(7)又は(8)で変形したもの;
    end
  else if データ並びの中に数値がある then
    begin (8)で変形したもの;
      history(新データ並び)
    end
  else if データ並びの中に数値がある then
    begin (10)又は(9)を適用;
      履歴ファイルに記録;
      新データ並び:=データ並びを(10)又は(9)で変形したもの;
    end
  else if (3)又は(4) が適用できる then
    begin (3)又は(4) を適用;
      履歴ファイルに記録;
      新データ並び:=データ並びを(3)又は(4)で変形したもの;
    end
  else if (5)又は(6) が適用できる then
    begin (5)又は(6) を適用;
      履歴ファイルに記録;
      新データ並び:=データ並びを(5)又は(6)で変形したもの;
    end
  else if (1)又は(2)を適用;
    begin (1)又は(2)を適用;
      履歴ファイルに記録;
      newデータ並び:=データ並びを(1)又は(2)で変形したもの;
    end
  else
    begin (5)と(6) を同時に適用;
      履歴ファイルに記録;
      新データ並び:=データ並びを(5)と(6)で変形したもの;
    end
  history(新データ並び)
end history

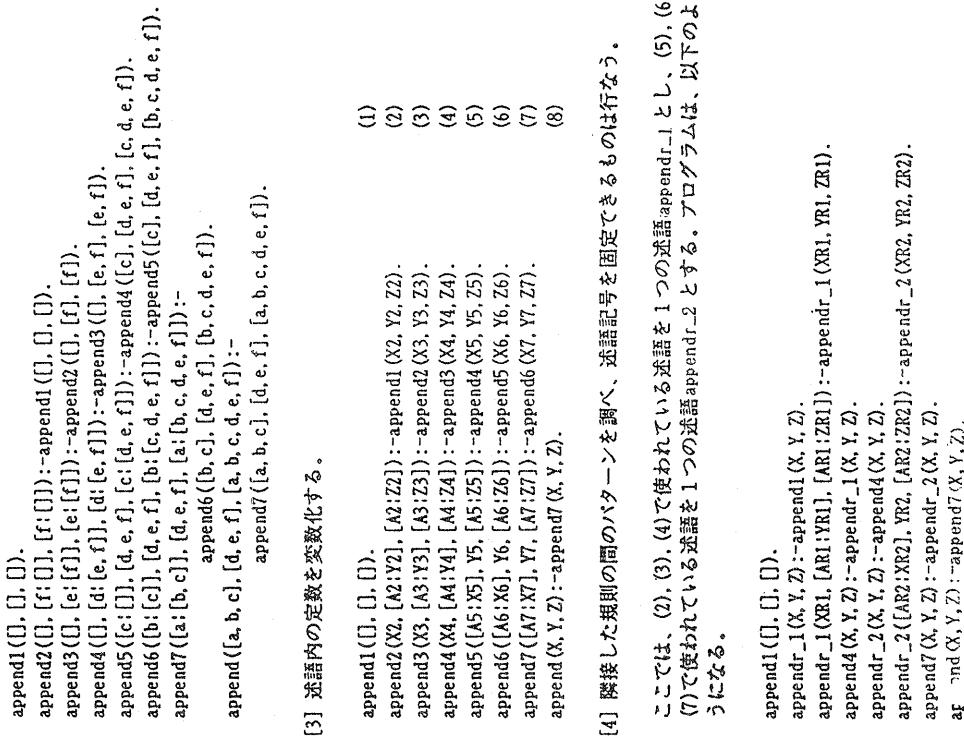
```

付録3 簡單で基本的な合成例

[1] 入力データを変形する規則が適用される。



[2] 入力データに対してのみ正しく働くプログラムを合成する。

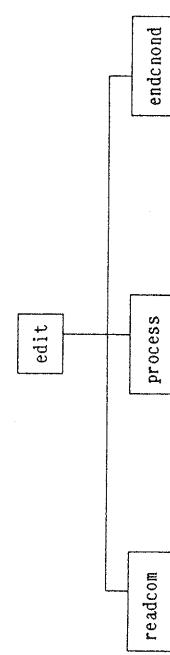


付録4. 本システムを使っての簡単な行エディタの作成

作成する行エディタの機能は、以下の通りである。

- (1) テキストの入出力
- (2) 行の削除
- (3) 行の挿入
- (4) 文字の削除
- (5) 文字の挿入
- (6) テキストの表示

おまかなかんプログラムの構成は以下の通りである。



本システムを用いて合成出来た部分の全体に対する割合を以下に示す。

| | 大きさ(バイト) | ステップ数(“.”の数) |
|---------------|----------|--------------|
| プログラム全体 | 2579 | 30 |
| 最下位の部分 | 1430 | 21 |
| 本システムで合成できた部分 | 910 | 16 |

下線の部分が主に本システムを用いて合成出来た部分である。

また、合成に用いた述語の動作例を以下に示す。

- (1) <delline : 1行削除 >
delline([[a, b, c, d, e], [f, g, h, i], [j, k, l, m, n, o], [p, q, r]], 3,
[[a, b, c, d, e], [f, g, h, i], [p, q, r]]).
- (2) <insline : 1行挿入 >
insline([[a, b, c, d], [e, f, g], [h, i, j, k, l, m], [n], [o, p, q, r]],
[[a, b, c, d], [e, f, g], [x, y, z], [h, i, j, k, l, m], [n], [o, p, q, r]]).

プログラムの大きさで考えると、プログラム全体に対して約35%、最下位のルーチン中で約6.4%のプログラムが本システムを用いて合成できな。またこれを、プログラムのステップ数に注目して考えると、プログラム全体に対して約5.3%、最下位のルーチン中で約7.6%のプログラムの合成に成功した。