# Algorithmisches Denken und imperative Programmierung

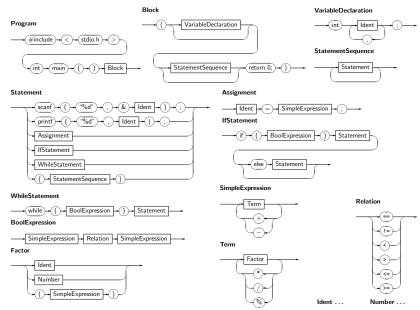
12. Vorlesung

Janis Voigtländer

Universität Bonn

Wintersemester 2012/13

## Wiederholung — C<sub>0</sub>



## Wiederholung — AM

$$AM = BZ \times DK \times HS \times Inp \times Out, mit:$$

 $\begin{array}{lll} \mathsf{BZ} &= \mathbb{N} & \mathsf{Befehlsz\ddot{a}hler} \\ \mathsf{DK} &= \mathbb{Z}^* & \mathsf{Datenkeller} \\ \mathsf{HS} &= \{h \mid h : \mathbb{N}_+ -\!\!\!\! \rightarrow \mathbb{Z}\} & \mathsf{Hauptspeicher} \\ \mathsf{Inp} &= \mathbb{Z}^* & \mathsf{Eingabeband} \\ \mathsf{Out} &= \mathbb{Z}^* & \mathsf{Ausgabeband} \end{array}$ 

- ▶ READ n: Lesen von Eingabeband in Hauptspeicher
- ▶ WRITE *n*: Ausgabe aus Hauptspeicher auf Ausgabeband
- ▶ LOAD n: Ablage aus Hauptspeicher auf Datenkeller
- STORE n: Entnahme aus Datenkeller in Hauptspeicher
- LIT z: Ablage einer Konstante auf Datenkeller
- ► ADD, MUL, SUB, DIV, MOD, LT, EQ, NE, GT, LE, GE: Berechnungen und Vergleiche (auf Datenkeller)
- ▶ JMP *n*: Sprung
- JMC n: bedingter Sprung abhängig von Datenkeller

# Übersetzung von C<sub>0</sub> nach AM — Warum?

```
#include <stdio.h>
int main()
{ int i,n,s;
  scanf("%d",&n);
  i = 1:
  s=0:
                                                       8: LE:
                                                                       15: STORE 3:
  while (i \le n)
                                     2: LIT 1:
                                                     9: JMC 21:
                                                                       16: LOAD 1:
     { s=s+i*i:
                                     3: STORE 1:
                                                     10: LOAD 3;
                                                                       17: LIT 1;
       i = i + 1:
                                    4: LIT 0:
                                                     11: LOAD 1;
                                                                       18: ADD;
                                    5: STORE 3;
                                                      12: LOAD 1;
                                                                       19: STORE 1;
  printf("%d",s);
                                    6: LOAD 1:
                                                      13: MUL:
                                                                       20: JMP 6:
  return 0:
                                    7: LOAD 2:
                                                      14: ADD:
                                                                       21: WRITE 3:
```

### Warum wollen wir von C<sub>0</sub> nach AM?

- ► C<sub>0</sub> auf der abstrakten Maschine ausführbar machen (sozusagen ein Compiler)
- ▶ C<sub>0</sub> eine formale Semantik geben

# Eingabesprache für die Übersetzung

Bezeichne  $W(\langle \mathsf{Program} \rangle)$  die Menge aller "Worte", die mit dem Startdiagramm **Program** beginnend erzeugt werden können. (analog auch für andere Teile des Syntaxdiagramm-Systems)

Unsere Eingabesprache sind alle Worte der Sprache  $W(\langle Program \rangle)$ , die folgende zwei (kontextsensitive) Nebenbedingungen erfüllen:

- 1. Jeder Bezeichner darf höchstens einmal in der eventuell vorhandenen VariableDeclaration deklariert sein.
- Wenn ein Bezeichner in der eventuell vorhandenen StatementSequence des Blocks auftritt, so muss er in der (dann unbedingt vorhandenen) VariableDeclaration des Blocks deklariert sein.

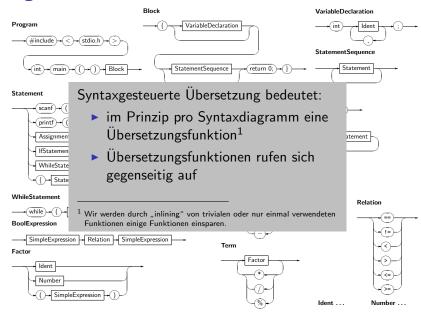
## Übersetzungsstrategie

Übersetzung erfolgt syntaxgesteuert, das heißt:

- ▶ Übersetzung von  $W(\langle \mathsf{Program} \rangle)$  benutzt Übersetzung von  $W(\langle \mathsf{Block} \rangle)$ .
- ▶ Übersetzung von  $W(\langle \mathsf{Block} \rangle)$  benutzt Analyse von  $W(\langle \mathsf{VariableDeclaration} \rangle)$  und Übersetzung von  $W(\langle \mathsf{StatementSequence} \rangle)$ .
- ▶ Übersetzung von  $W(\langle StatementSequence \rangle)$  benutzt Übersetzung von  $W(\langle Statement \rangle)$ .

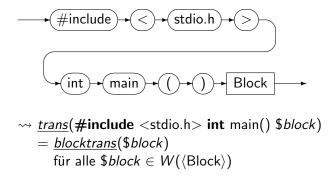
**•** 

## Auf geht's, aber wie?



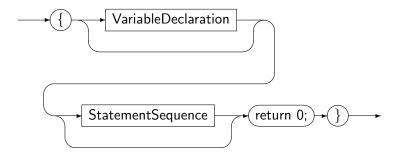
# Syntaxgesteuerte Übersetzung (I)

## Program



# Syntaxgesteuerte Übersetzung (II)

#### **Block**



```
 \overset{blocktrans}{(\$vardecl \$statseq return 0;\})} = \underbrace{stseqtrans}(\$statseq, \cdots)  für alle \$vardecl \in \{\varepsilon\} \cup W(\langle VariableDeclaration \rangle)  und \$statseq \in \{\varepsilon\} \cup W(\langle StatementSequence \rangle)
```

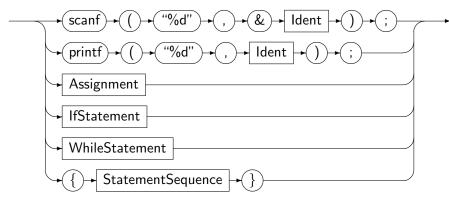
# Syntaxgesteuerte Übersetzung (III)

## **StatementSequence**

```
Statement \\ \hookrightarrow \underline{stseqtrans}(\$stat_1 \dots \$stat_n, \cdots) \\ = \underline{sttrans}(\$stat_1, \cdots) \\ \vdots \\ \underline{sttrans}(\$stat_n, \cdots) \\ \text{für alle } \$stat_1, \dots, \$stat_n \in W(\langle \mathsf{Statement} \rangle)
```

# Syntaxgesteuerte Übersetzung (IV)

### Statement

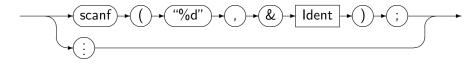


→ Fallunterscheidung,

```
zum Beispiel: \underline{sttrans}(scanf("%d",&$id);,\cdots)
= READ ?
für alle \$id \in W(\langle Ident \rangle)
```

# Syntaxgesteuerte Übersetzung (IV)

#### Statement

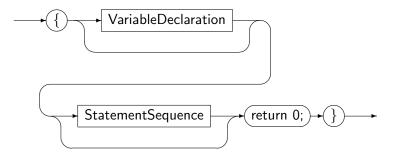


```
ightharpoonup Fallunterscheidung, zum Beispiel: \underline{sttrans}(\text{scanf}(\text{"%d",&$id});,\cdots) = READ ? für alle \text{$id \in W(\langle \text{Ident} \rangle)$}
```

Wir brauchen Informationen über die Zuordnung von Bezeichnern (im Programm) zu Speicherplätzen (im HS der AM)!

## Erzeugung einer sogenannten Symboltabelle (I)

### **Block**



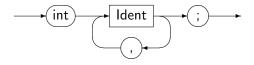
```
 \overset{blocktrans}{(\$vardecl \$statseq return 0;\})} = \underbrace{stseqtrans}(\$statseq, \underbrace{mksymtab}(\$vardecl), \cdots)  für alle \$vardecl \in \{\varepsilon\} \cup W(\langle Variable Declaration \rangle)  und \$statseq \in \{\varepsilon\} \cup W(\langle Statement Sequence \rangle)
```

Menge der Symboltabellen:

$$\mathsf{Tab} = \{ \mathsf{tab} \mid \mathsf{tab} : W(\langle \mathsf{Ident} \rangle) \longrightarrow \mathbb{N}_+ \}$$

## Erzeugung einer sogenannten Symboltabelle (II)

#### **VariableDeclaration**



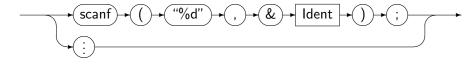
```
ightarrow \underline{mksymtab}(arepsilon) = [] 	ext{ (leere Abbildung)} 
\underline{mksymtab}(\mathbf{int} \ \$id_1, \dots, \$id_m;) 
= [\$id_1/1, \dots, \$id_m/m] 
\text{ für alle } \$id_1, \dots, \$id_m \in W(\langle \mathsf{Ident} \rangle)
```

Die Symboltabelle wird von <u>stseqtrans</u> aus in weitere Übersetzungsfunktionen propagiert!

Zuordnung in der Symboltabelle ist eindeutig, wegen der ersten kontextsensitiven Nebenbedingung (keine Doppeldeklarationen).

# Syntaxgesteuerte Übersetzung (IV)

#### Statement



 $\sim$  Fallunterscheidung, zum Beispiel:  $\underline{sttrans}(\text{scanf}(\text{"%d",&$id$});,\underline{tab},\cdots)$  = wenn  $\underline{tab}(\text{$id$}) = n$ , dann READ n; für alle  $\text{$id \in W(\langle \text{Ident} \rangle)}$  und  $tab \in \text{Tab}$ 

Zugriff auf *tab* hier ist immer definiert, wegen der zweiten kontextsensitiven Nebenbedingung (nur deklarierte Bezeichner dürfen verwendet werden).

# Syntaxgesteuerte Übersetzung (V)

## **Assignment**

Geht auf, wenn (weil!) <u>simpleexptrans</u>(\$exp, tab) zu einem Stack mit dem Berechnungsergebnis an oberster Position führt (wobei nicht tiefer in den Datenkeller eingegriffen wird).

## Einschub: Prinzip der Berechnungsübersetzung

Jede Rechnung in  $C_0$  wird in AM mittels der sogenannten "reverse polish notation" (Postfix-Notation) umgesetzt, d.h. Operatoren stehen hinter den Operanden.

Beispiel: 
$$1 + 2 \Rightarrow 12 +$$

Der Vorteil ist, dass keine Klammerung mehr nötig ist, da jeder Operator nur so viel Operanden konsumiert, wie er benötigt.

Beispiel: 
$$2 * (1 + 3 - 2) \Rightarrow 2 \ 1 \ 3 + 2 - *$$

Legt man Operanden auf den Stack und führt Operatoren jeweils direkt aus, dann ergeben sich die Zwischen- und Endergebnisse.

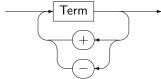
## Beispiel:

$$2*(1+3-2) \Rightarrow 2 \cdot 1 \cdot 3 + \Rightarrow 2 \cdot 4 \Rightarrow 2 \cdot 4 \cdot 2 - \Rightarrow 2 \cdot 2 \Rightarrow 2 \cdot 2 \Rightarrow 4$$

Wir wissen nun auch: Jede Berechnung nimmt nie mehr vom Stack als sie drauflegt.

# Syntaxgesteuerte Übersetzung (VI)

### SimpleExpression



```
\rightsquigarrow simpleexptrans(t_1 \cdot p_2 \cdot t_2 \dots \cdot p_n \cdot t_n, tab)
    = termtrans(\$t_1, tab)
        termtrans(\$t_2, tab)
        OP<sub>2</sub>:
        termtrans(\$t_n, tab)
        OP_n:
        für alle t_1, \ldots, t_n \in W(\langle \text{Term} \rangle),
        p_2, \dots, p_n \in \{+, -\} \text{ und } tab \in \mathsf{Tab},
        wobei OP_i = ADD, falls p_i = +
                 OP_i = SUB, falls p_i = -
```

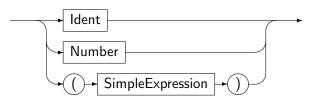
# Syntaxgesteuerte Übersetzung (VII)

```
Factor *
```

```
\rightarrow termtrans(f_1 \otimes p_2 \otimes f_2 \dots \otimes p_n \otimes f_n, tab)
     = factortrans(\$f_1, tab)
         factortrans(\$f_2, tab)
         OP<sub>2</sub>;
         factortrans(\$f_n, tab)
         OP<sub>n</sub>:
         für alle f_1, \ldots, f_n \in W(\langle Factor \rangle),
         p_2, \dots, p_n \in \{*,/,\%\} und p_n \in \{*,d\} und p_n \in \{*,d\}
         wobei OP_i = MUL, falls p_i = *
                  OP_i = DIV, falls p_i = /
                  OP_i = MOD, falls p_i = \%
```

# Syntaxgesteuerte Übersetzung (VIII)

### **Factor**



- - = wenn tab(\$id) = n, dann LOAD n; für alle  $\$id \in W(\langle Ident \rangle)$  und  $tab \in Tab$

## $\underline{factortrans}(\$z, tab)$

- = LIT z; für alle  $z \in W(\langle Number \rangle)$  und  $tab \in Tab$
- $\underline{factortrans}((\$se), tab)$
- $= \underline{simpleexptrans}(\$se, tab)$ für alle  $\$se \in W(\langle SimpleExpression \rangle)$  und  $tab \in Tab$

# Syntaxgesteuerte Übersetzung (IX)

### BoolExpression

```
→ SimpleExpression |→ Relation |
                                            → SimpleExpression
\rightsquigarrow boolexptrans(\$se_1 \$rel \$se_2, tab)
   = simple exptrans(\$se_1, tab)
      simple exptrans(\$se_2, tab)
      REL:
      für alle \$se_1,\$se_2 \in W(\langle SimpleExpression \rangle),
      rel \in \{==,!=,<,>,<=,>=\} \text{ und } tab \in Tab,
      wobei REL = EQ. falls rel = =
              REL = NE. falls rel = !=
              REL = LT. falls rel = <
```

# Syntaxgesteuerte Übersetzung (X)

### WhileStatement

```
BoolExpression
                                                     Statement
\rightsquigarrow sttrans(while ($exp) $stat, tab, \cdots)
    = boolexptrans($exp, tab)
       JMC ?:
       sttrans(\$stat, tab, \cdots)
       JMP ?:
       für alle \$exp \in W(\langle BoolExpression \rangle),
      \$stat \in W(\langle \mathsf{Statement} \rangle) und tab \in \mathsf{Tab}
                  keine konkreten Adressen bekannt.
     Problem:
                  hängen unter anderem von Länge des übersetzten
                     Codes für $exp und $stat ab
      Lösung:
                  zunächst nur abstrakte Adressen, später
                     Nachbearbeitung
                  "baumstrukturierte Adressen": Listen über
                     natürlichen Zahlen (Notation 3.2.4.1)
```

# Syntaxgesteuerte Übersetzung (X)

# Syntaxgesteuerte Übersetzung (XI)

```
blocktrans({$vardecl $statseg return 0;})
= stseqtrans($statseq, mksymtab($vardecl), 1)
   für alle vardecl \in \{\varepsilon\} \cup W(\langle Variable Declaration \rangle)
   und \$statseq \in \{\varepsilon\} \cup W(\langle StatementSequence \rangle)
stseqtrans(\$stat_1 \ldots \$stat_n, tab, a)
= sttrans(\$stat_1, tab, a.1)
   sttrans(\$stat_n, tab, a.n)
   für alle \$stat_1, \ldots, \$stat_n \in W(\langle Statement \rangle), tab \in Tab und <math>a \in \mathbb{N}^*
```

# Syntaxgesteuerte Übersetzung (XII)

```
sttrans(if ($exp) $stat, tab, a)
= boolexptrans($exp, tab)
       JMC a:
       sttrans($stat, tab, a.1)
   a:
   für alle \$exp \in W(\langle BoolExpression \rangle), \$stat \in W(\langle Statement \rangle),
   tab \in \mathsf{Tab} \ \mathsf{und} \ a \in \mathbb{N}^*
sttrans(if ($exp) $stat_1 else $stat_2, tab, a)
= boolexptrans($exp, tab)
         JMC a:
         sttrans(\$stat_1, tab, a.1)
         JMP a.3:
      a: sttrans($stat<sub>2</sub>, tab, a.2)
   a.3:
   für alle \$exp \in W(\langle BoolExpression \rangle),
   \$stat<sub>1</sub>, \$stat<sub>2</sub> \in W(\langle Statement \rangle), tab \in Tab und a \in \mathbb{N}^*
```

# Syntaxgesteuerte Übersetzung (XIII)



# **Zusammenfassung (I)**

```
trans(#include <stdio.h> int main() $block)
= blocktrans($block)
blocktrans({$vardecl $statseg return 0;})
= stseqtrans($statseq, mksymtab($vardecl), 1)
mksymtab(\varepsilon) = []
mksymtab(int $id_1, ..., $id_m;) = [$id_1/1, ..., $id_m/m]
stseqtrans(\$stat_1 \ldots \$stat_n, tab, a)
= sttrans(\$stat_1, tab, a.1)
  sttrans(\$stat_n, tab, a.n)
```

# Zusammenfassung (II)

```
sttrans(\$id = \$exp;, tab, a)
= wenn tab(\$id) = n, dann:
  simpleexptrans($exp, tab)
  STORE n:
sttrans(if ($exp) $stat, tab, a)
= boolexptrans($exp, tab)
     JMC a:
     sttrans($stat, tab, a.1)
  a:
sttrans(if ($exp) $stat_1 else $stat_2, tab, a)
   <u>boolexptrans(</u>$exp, tab)
       JMC a:
       sttrans(\$stat_1, tab, a.1)
       JMP a.3:
    a: sttrans($stat2, tab, a.2)
  a.3:
```

# Zusammenfassung (III)

```
sttrans(while ($exp) $stat, tab, a)
= a.2: boolexptrans($exp, tab)
       JMC a:
       sttrans($stat, tab, a.1)
       JMP a.2:
    a:
sttrans(scanf("%d",&$id);, tab, a)
= wenn tab(\$id) = n, dann READ n;
sttrans(printf ("%d",$id);, tab, a)
= wenn tab(\$id) = n, dann WRITE n;
\underline{sttrans}(\{\$stat_1 \ldots \$stat_n\}, tab, a)
= stseqtrans(\$stat_1 ... \$stat_n, tab, a)
```

# **Zusammenfassung (IV)**

```
boolexptrans($se<sub>1</sub> $rel $se<sub>2</sub>, tab)
= simple exptrans(\$se_1, tab)
   simpleexptrans($se2, tab)
   REL:
   wobei REL = EQ, falls rel = =
simple exptrans( t_1 p_2 t_2 ... p_n t_n, tab)
= termt_{\underline{rans}}(\$t_1, tab)
   termtrans(\$t_2, tab)
   OP_2:
   termtrans(\$t_n, tab)
   OP<sub>n</sub>:
   wobei OP_i = ADD, falls p_i = +
          OP_i = SUB, falls p_i = -
```

# **Zusammenfassung (V)**

```
\underline{termtrans}(\$f_1 \$op_2 \$f_2 \dots \$op_n \$f_n, tab)
= factortrans(\$f_1, tab)
   factortrans(\$f_2, tab)
   OP<sub>2</sub>:
   factortrans(\$f_n, tab)
   \mathsf{OP}_n:
   wobei OP_i = MUL, falls p_i = *
factortrans($id, tab)
= wenn tab(\$id) = n, dann LOAD n;
factortrans(\$z, tab) = LIT \$z;
\underline{factortrans}((\$se), tab) = \underline{simpleexptrans}(\$se, tab)
```

## **Zusammenfassung (VI)**

## C<sub>0</sub> wird auf AM abgebildet, indem:

- "atomare Befehle" (wie scanf oder printf) direkt umgesetzt/übersetzt werden,
- Kontrollstrukturen in (geeignet arrangierte) Sprünge übersetzt werden, wobei
  - wegen syntaktischer Schachtelung (Blockstruktur, komplexe Statements als Teile anderer Statements) die Übersetzung in flach strukturierten AM-Code nicht (bzw. schwer) möglich ist, daher ein Umweg über baumstrukturierte Adressen gegangen wird,
- Schachtelung bei Ausdrücken mittels Stackprinzip umgesetzt wird.

schließlich ...

```
#include < stdio . h>
int main()
{ int i, n, s;
  scanf("%d",&n);
  i = 1;
  s=0:
  while (i \le n)
    \{ s=s+i*i;
     i = i + 1;
   printf("%d",s);
  return 0;
```

```
trans(#include <stdio.h> int main() { ... return 0; })
= blocktrans(\{ int i,n,s; scanf("%d",&n); ... return 0; \})
= stseqtrans(scanf("%d",&n); ... printf("%d",s);,
       mksymtab(int i,n,s;), 1)
= \underline{stseqtrans}(scanf("%d",&n); \dots printf("%d",s);,
       [i/1, n/2, s/3], 1)
= sttrans(scanf("%d",&n); , tab_1, 1.1)
  sttrans(i=1; tab_1, 1.2)
  sttrans(s=0; tab_1, 1.3)
  sttrans(while (i \le n) { s=s+i*i; i=i+1; }, tab_1, 1.4)
  sttrans(printf ("%d",s); , tab_1, 1.5)
```

```
= \begin{array}{ll} \text{READ 2;} \\ \underline{simpleexptrans}(1,tab_1) & \text{STORE 1;} \\ \underline{simpleexptrans}(0,tab_1) & \text{STORE 3;} \\ 1.4.2: \underline{boolexptrans}(i\!<\!=\!n\,,tab_1) \\ \text{JMC 1.4;} \\ \underline{sttrans}(\{\text{ s=s+i*i; i=i+1; }\},tab_1\,,1.4.1) \\ \text{JMP 1.4.2;} \\ 1.4: \text{WRITE 3;} \end{array}
```

```
= \qquad \text{READ 2;} \\ \text{LIT 1;} \qquad \text{STORE 1;} \\ \text{LIT 0;} \qquad \text{STORE 3;} \\ 1.4.2: \underbrace{\textit{termtrans}}_{\text{JMC 1.4;}} (i, tab_1) \quad \underbrace{\textit{termtrans}}_{\text{termtrans}} (n, tab_1) \quad \text{LE;} \\ \text{JMC 1.4;} \\ \underbrace{\textit{simpleexptrans}}_{\text{simpleexptrans}} (i+i, tab_1) \quad \text{STORE 3;} \\ \underbrace{\textit{simpleexptrans}}_{\text{JMP 1.4.2;}} (i+1, tab_1) \quad \text{STORE 1;} \\ \text{JMP 1.4.2;} \\ 1.4: \text{WRITE 3;} \\ \end{aligned}
```

```
= READ 2;
LIT 1; STORE 1;
LIT 0; STORE 3;
1.4.2 : \underline{factortrans}(i, tab_1) \underline{factortrans}(n, tab_1) LE;
JMC 1.4;
\underline{termtrans}(s, tab_1) \underline{termtrans}(i*i, tab_1)
ADD; STORE 3;
\underline{termtrans}(i, tab_1) \underline{termtrans}(1, tab_1) ADD; STORE 1;
JMP 1.4.2;
1.4 : WRITE 3;
```

```
= READ 2;

LIT 1; STORE 1;

LIT 0; STORE 3;

1.4.2: LOAD 1; LOAD 2; LE;

JMC 1.4;

<u>factortrans(s, tab_1)</u> <u>factortrans(i, tab_1)</u> <u>factortrans(i, tab_1)</u>

MUL; ADD; STORE 3;

<u>factortrans(i, tab_1)</u> <u>factortrans(1, tab_1)</u> ADD; STORE 1;

JMP 1.4.2;

1.4: WRITE 3;
```

```
= READ 2;
LIT 1; STORE 1;
LIT 0; STORE 3;
1.4.2: LOAD 1; LOAD 2; LE;
JMC 1.4;
LOAD 3; LOAD 1; LOAD 1;
MUL; ADD; STORE 3;
LOAD 1; LIT 1; ADD; STORE 1;
JMP 1.4.2;
1.4: WRITE 3;
```

## Beispiel — übersetzt

```
STORE 3:
     READ 2:
                          LE:
     LIT 1:
                          JMC 1.4:
                                              LOAD 1:
     STORE 1:
                          LOAD 3;
                                              LIT 1:
     LIT 0;
                          LOAD 1:
                                              ADD:
     STORE 3;
                          LOAD 1:
                                              STORE 1:
1.4.2: LOAD 1:
                          MUL:
                                              JMP 1.4.2;
     LOAD 2:
                          ADD:
                                          1.4: WRITE 3:
```

## Linearisierung:

- 1. Durchnummerierung der Befehle, beginnend mit 1
- 2. Merken von Paaren aus baumstrukturierter Adresse und nummerierter Adresse
- Anpassen von Sprungbefehlen entsprechend der gemerkten Paare

## Beispiel — linearisiert

```
1: READ 2;
                          15: STORE 3;
             8: LE;
2: LIT 1;
             9: JMC 21;
                         16: LOAD 1;
3: STORE 1;
            10: LOAD 3;
                         17: LIT 1:
4: LIT 0;
            11: LOAD 1;
                         18: ADD;
            12: LOAD 1;
5: STORE 3;
                         19: STORE 1:
6: LOAD 1;
            13: MUL; 20: JMP 6;
7: LOAD 2;
            14: ADD; 21: WRITE 3;
```