

ALBERT LUDWIGS UNIVERSITÄT FREIBURG

TECHNISCHE FAKULTÄT

PicoC-Compiler

Übersetzung einer Untermenge von C in den Befehlssatz der RETI-CPU

BACHELORARBEIT

Abgabedatum: 28th April 2022

Author:
Jürgen Mattheis

Gutachter:
Prof. Dr. Scholl

Betreuung:
M.Sc. Seufert

Eine Bachelorarbeit am Lehrstuhl für
Betriebssysteme

ERKLÄRUNG

Hiermit erkläre ich, dass ich diese Abschlussarbeit selbständig verfasst habe, keine anderen als die angegebenen Quellen/Hilfsmittel verwendet habe und alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen wurden, als solche kenntlich gemacht habe. Darüber hinaus erkläre ich, dass diese Abschlussarbeit nicht, auch nicht auszugsweise, bereits für eine andere Prüfung angefertigt wurde.

Inhaltsverzeichnis

Abbildungsverzeichnis	I
Codeverzeichnis	II
Tabellenverzeichnis	III
Definitionsverzeichnis	IV
Grammatikverzeichnis	V
1 Implementierung	1
1.1 Lexikalische Analyse	3
1.1.1 Konkrete Syntax für die Lexikalische Analyse	3
1.1.2 Codebeispiel	5
1.2 Syntaktische Analyse	5
1.2.1 Umsetzung von Präzidenz und Assoziativität	5
1.2.2 Konkrete Syntax für die Syntaktische Analyse	10
1.2.3 Ableitungsbaum Generierung	12
1.2.3.1 Codebeispiel	14
1.2.3.2 Ausgabe des Ableitungsbaums	15
1.2.4 Ableitungsbaum Vereinfachung	15
1.2.4.1 Codebeispiel	17
1.2.5 Abstrakt Syntax Tree Generierung	18
1.2.5.1 PicoC-Knoten	20
1.2.5.2 RETI-Knoten	25
1.2.5.3 Kompositionen von PicoC-Knoten und RETI-Knoten mit besonderer Bedeutung	26
1.2.5.4 Abstrakte Syntax	28
1.2.5.5 Codebeispiel	29
1.2.5.6 Ausgabe des Abstrakter Syntaxbaum	30
1.3 Code Generierung	30
1.3.1 Passes	32
1.3.1.1 PicoC-Shrink Pass	33
1.3.1.1.1 Aufgabe	33
1.3.1.1.2 Abstrakte Syntax	33
1.3.1.1.3 Codebeispiel	34
1.3.1.2 PicoC-Blocks Pass	36
1.3.1.2.1 Aufgabe	36
1.3.1.2.2 Abstrakte Syntax	36
1.3.1.2.3 Codebeispiel	38
1.3.1.3 PicoC-ANF Pass	39
1.3.1.3.1 Aufgabe	39
1.3.1.3.2 Abstrakte Syntax	40
1.3.1.3.3 Codebeispiel	42
1.3.1.4 RETI-Blocks Pass	43
1.3.1.4.1 Aufgabe	43
1.3.1.4.2 Abstrakte Syntax	43
1.3.1.4.3 Codebeispiel	44

1.3.1.5	RETI-Patch Pass	47
1.3.1.5.1	Aufgabe	47
1.3.1.5.2	Abstrakte Syntax	47
1.3.1.5.3	Codebeispiel	48
1.3.1.6	RETI Pass	51
1.3.1.6.1	Aufgabe	51
1.3.1.6.2	Konkrete und Abstrakte Syntax	51
1.3.1.6.3	Codebeispiel	53

Literatur	A
------------------	----------

Abbildungsverzeichnis

1.1	Ableitungsbäume zu den beiden Ableitungen	7
1.2	Ableitungsbaum nach Parsen eines Ausdrucks	16
1.3	Ableitungsbaum nach Vereinfachung	17
1.4	Abstrakter Syntaxbaum Generierung ohne Umdrehen	19
1.5	Abstrakter Syntaxbaum Generierung mit Umdrehen	19
1.6	Cross-Compiler Kompiliervorgang ausgeschrieben	31
1.7	Cross-Compiler Kompiliervorgang Kurzform	31
1.8	Architektur mit allen Passes ausgeschrieben	32

Codeverzeichnis

1.1	PicoC-Code des Codebeispiels	5
1.2	Tokens für das Codebeispiel	5
1.3	Ableitungsbaum nach Ableitungsbaum Generierung	14
1.4	Ableitungsbaum nach Ableitungsbaum Vereinfachung	18
1.5	Aus vereinfachtem Ableitungsbaum generierter Abstrakter Syntaxbaum	29
1.6	PicoC Code für Codebeispiel	35
1.7	Abstrakter Syntaxbaum für Codebeispiel	36
1.8	PicoC-Blocks Pass für Codebeispiel	39
1.9	PicoC-ANF Pass für Codebeispiel	43
1.10	RETI-Blocks Pass für Codebeispiel	47
1.11	RETI-Patch Pass für Codebeispiel	51
1.12	RETI Pass für Codebeispiel	55

Tabellenverzeichnis

1.1	Präzidenzregeln von PicoC	6
1.2	Zuordnung der Bezeichnungen von Produktionsregeln zu Operatoren	8
1.3	PicoC-Knoten Teil 1	21
1.4	PicoC-Knoten Teil 2	22
1.5	PicoC-Knoten Teil 3	23
1.6	PicoC-Knoten Teil 4	24
1.7	RETI-Knoten	26
1.8	Kompositionen von PicoC-Knoten und RETI-Knoten mit besonderer Bedeutung	27

Definitionsverzeichnis

1.1	Metasyntax	1
1.2	Metasprache	1
1.3	Erweiterte Backus-Naur-Form (EBNF)	1
1.4	Dialekt der EBNF aus Lark	2
1.5	Abstrakte Syntax Form (ASF)	2
1.6	Earley Parser	12
1.7	Earley Recognizer	12
1.8	Label	25
1.9	Token-Knoten	25
1.10	Container-Knoten	25
1.11	Symboltabelle	40

Grammatikverzeichnis

1.1.1 Grammatik der Konkreten Syntax der Sprache L_{PicoC} für die Lexikalische Analyse in EBNF	4
1.2.1 Undurchdachte Konkrete Syntax der Sprache L_{PicoC} für die Syntaktische Analyse in EBNF	6
1.2.2 Durchdachte Konkrete Syntax der Sprache L_{PicoC} für die Syntaktische Analyse in EBNF . .	7
1.2.3 Beispiel für eine unäre rechtsassoziative Produktion	8
1.2.4 Beispiel für eine unäre linksassoziative Produktion	8
1.2.5 Beispiel für eine linksassoziative Produktion	9
1.2.6 Beispiel für eine linksassoziative Produktion	9
1.2.7 Durchdachte Konkrete Syntax für Operatorpräzedenz in EBNF	10
1.2.8 Grammatik der Konkreten Syntax der Sprache L_{PicoC} für die Syntaktische Analyse in EBNF, Teil 1	11
1.2.9 Grammatik der Konkreten Syntax der Sprache L_{PicoC} für die Syntaktische Analyse in EBNF, Teil 2	12
1.2.10 Abstrakte Syntax der Sprache L_{PioC}	28
1.3.1 Abstrakte Syntax der Sprache L_{PioC_Shrink}	34
1.3.2 Abstrakte Syntax der Sprache L_{PioC_Blocks}	37
1.3.3 Abstrakte Syntax der Sprache L_{PioC_ANF}	41
1.3.4 Abstrakte Syntax der Sprache L_{RETI_Blocks}	44
1.3.5 Abstrakte Syntax der Sprache L_{RETI_Patch}	48
1.3.6 Konkrete Syntax der Sprache L_{RETI} für die Lexikalische Analyse in EBNF	52
1.3.7 Konkrete Syntax der Sprache L_{RETI} für die Syntaktische Analyse in EBNF	52
1.3.8 Abstrakte Syntax der Sprache L_{RETI}	52

1 Implementierung

In diesem Kapitel wird, nachdem im Kapitel ?? die nötigen **theoretischen Grundlagen** des **Compilerbau** vermittelt wurden, nun auf die **Implementierung** des **PicoC-Compilers** eingegangen. Aufgeteilt in die selben Kategorien **Lexikalische Analyse 1.1**, **Syntaktische Analyse 1.2** und **Code Generierung 1.3**, wie in Kapitel ??, werden in den folgenden Unterkapiteln die einzelnen **Zwischenschritte** vom einem **Programm** in der **Konkreten Syntax** der Sprache L_{PicoC} hin zum einem Programm mit derselben **Semantik** in der **Konkreten Syntax** der Sprache L_{RETI} erklärt.

Für das Parsen¹ des Programmes in der **Konkreten Syntax** der Sprache L_{PicoC} wird das **Lark Parsing Toolkit**^{2 3} verwendet. Das **Lark Parsing Toolkit** ist eine Bibliothek, die es ermöglicht mittels eines in einem **eigenen Dialekt** der **Erweiterten Back-Naur-Form** (Definition 1.3 bzw. für den Dialekt von Lark Definition 1.4) spezifizierten Grammatik der **Konkreten Syntax** ein Programm in ebendieser **Konkreten Syntax** zu parsen und daraus einen **Ableitungsbaum** für die kompilerintere Weiterverarbeitung zu generieren.

Definition 1.1: Metasyntax

*Steht für den **Aufbau** einer **Metasprache** (Definition 1.2), der durch eine **Grammatik** oder **Natürliche Sprache** beschrieben werden kann.*

Definition 1.2: Metasprache

*Eine Metasprache ist eine **Sprache**, die dazu genutzt wird **andere Sprachen** zu **beschreiben**^a.*

^aDas „Meta“ drückt allgemein aus, dass sich etwas auf einer **höheren Ebene** befindet. Um über die Ebene sprechen zu können, in der man sich **selbst befindet**, muss man von einer **höheren, außenstehenden Ebene** darüber reden.

Definition 1.3: Erweiterte Backus-Naur-Form (EBNF)

*Die Erweiterte Backus-Naur-Form^a ist eine **Metasyntax** (Definition 1.1) die dazu verwendet wird **Kontextfreier Grammatiken** darzustellen.^{b c}*

*Die Erweiterte Backus-Naur-Form ist zwar **standartisiert** und die Spezifikation des Standards kann unter **Link**^d aufgefunden werden, allerdings werden in der Praxis, wie z.B. in Lark oft **eigene Notationen** verwendet.*

^aDer Name kommt daher, dass es eine **Erweiterung** der **Backus-Naur-Form** ist, die hier allerdings **nicht** weiter erläutert wird.

^bNebel, „Theoretische Informatik“.

^cGrammar Reference — Lark documentation.

^d<https://standards.iso.org/ittf/PubliclyAvailableStandards/>.

¹Wobei beim **Parsen** auch das **Lexen** inbegriffen ist.

²Lark - a parsing toolkit for Python.

³Shinan, lark.

Definition 1.4: Dialekt der EBNF aus Lark

Das **Lark Parsing Toolkit** verwendet eine *eigene Notation* für die **Erweiterte Backus-Naur-Form**, die sich teilweise in einzelnen Aspekten von der Syntax aus dem **Standard** unterscheidet und unter *Link^a* dokumentiert ist.

Ein für die Grammatiken des PicoC-Compilers wichtiger **Unterschied** ist z.B., dass dieser Dialekt anstelle von **geschweiften Klammern** `{}` für die Darstellung von **Wiederholung**, den aus **regulären Ausdrücken** bekannten ***-Quantor optional** zusammen mit **runden Klammern** `()` verwendet: `()*`.^b

^a<https://lark-parser.readthedocs.io/en/latest/grammar.html>.

^bBzw. kann der ***-Quantor** auch **keinmal** wiederholen bedeuten.

Das **Lark Parsing Toolkit** wurde vor allem deswegen gewählt, weil es sehr **einfach in der Verwendung** ist. Andere derartige Tools, wie z.B. **ANTLR⁴** sind **Parser Generatoren**, die zur **Konkreten Syntax** einer Sprache einen **Parser** in einer vorher bestimmten Programmiersprache generieren, anstatt wie das **Lark Parsing Toolkit** bei Angabe einer **Konkreten Syntax** direkt ein Programm in dieser Konkreten Syntax **parsen** und einen **Ableitungsbaum** dafür generieren zu können.

Eine möglichst **geringe Laufzeit** durch Verwenden der effizientesten Algorithmen zu erreichen war keine der Hauptzielsetzungen für den **PicoC-Compiler**, da der **PicoC-Compiler** vor allem als **Lerntool** konzipiert ist, mit dem Studenten lernen können, wie der **Kompiliervorgang** von der Programmiersprache L_{PicoC} zur Maschinensprache L_{RETI} funktioniert. Eine ausführliche Diskussion zur Priorisierung Laufzeit wurde in Unterkapitel ?? geführt. Lark besitzt des Weiteren eine **sehr gute Dokumentation** *Welcome to Lark's documentation! — Lark documentation*, sodass anderen Studenten, die den **PicoC-Compiler** vielleicht in ihr Projekt einbinden wollen, unkompliziert **Erweiterungen** für den **PicoC-Compiler** schreiben können.

Neben den **Konkreten Syntaxen⁵**, die aufgrund der Verwendung des **Lark Parsing Toolkit** in einem **eigenen Dialekt** der **Erweiterter Back-Naur-Form** spezifiziert sind, werden in den folgenden Unterkapiteln die **Abstrakte Syntaxen**, welche spezifizieren, welche Kompositionen für die **Abstrakter Syntaxbaums** der verschiedenen **Passes** erlaubt sind in einer bewusst anderen Notation aufgeschrieben, die allerdings Ähnlichkeit mit dem Dialekt der **Erweiterten Backus-Naur-Form** aus dem **Lark Parsing Toolkit** hat.

Die Notation für die **Abstrakte Syntax** unterscheidet sich bewusst von der **Erweiterten Backus-Naur-Form**, da in der Abstrakten Syntax **Kompositionen von Knoten** beschrieben werden, die **klar auszumachen** sind, wodurch es die Grammatik nur **unnötig verkomplizieren** würde, wenn man die **Erweiterte Backus-Naur-Form** verwenden würde. Es gibt leider **keine** Standardnotation für die **Abstrakte Syntax**, die sich deutlich durchgesetzt hat, daher wird für die Abstrakte Syntaxen eine eigene **Abstract Syntax Form Notation** (Definition 1.5) verwendet. Des Weiteren trägt das Verwenden einer **unterschiedlichen Notation** für **Konkrete** und **Abstrakte Syntax** auch dazu bei, dass man beide direkter voneinander unterscheiden kann.

Definition 1.5: Abstrakte Syntax Form (ASF)

Die **Abstrakte Syntax Form** ist eine *eigene Metasyntax* für die **Grammatiken von Abstrakten Syntaxen**, die für diese Bachelorarbeit definiert wurde und sich von dem **Dialekt der Backus-Naur-Form** des **Lark Parsing Toolkit** nur dadurch unterscheidet, dass **Terminalsymbole** nicht von `"` eingeschlossen sein müssen, da die **Knoten** in der **Abstrakten Syntax**, sowieso schon **klar auszumachen** sind und von anderen Symbolen der **Metasprache** leicht zu unterscheiden sind.

Letztendlich geht es allerdings nur darum, dass aufgrund der Verwendung des **Lark Parsing Toolkit** die **Konkrete Syntax** in einem eigenen Dialekt der **Erweiterter Backus-Naur-Form** angegeben sein muss

⁴**ANTLR**.

⁵Der **Plural** von Syntax ist **Syntaxen**, wie es in Quelle *Syntax* verifiziert werden kann.

und für das Implementieren der Passes die **Abstrakte Syntax** für den **Programmierer** möglichst **einfach verständlich** sein sollte, weshalb sich die **Abstrakte Syntax Form** gut dafür eignet.

1.1 Lexikalische Analyse

Für die **Lexikalische Analyse** ist es nur notwendig eine Grammatik zu definieren, die den Teil der **Konkreten Syntax** beschreibt, der die **verschiedenen Pattern** für die verschiedenen Token der Sprache L_{PicoC} beschreibt, also den Teil der für die **Lexikalische Analyse** wichtig ist. Diese Grammatik wird dann vom **Lark Parsing Toolkit** dazu verwendet ein Programm in **Konkreter Syntax** zu lexen und daraus Tokens für die **Syntaktische Analyse** zu erstellen, wie es im Unterkapitel ?? erläutert ist.

1.1.1 Konkrete Syntax für die Lexikalische Analyse

In der Grammatik 1.1.1 für die **Lexikalische Analyse** stehen **großgeschriebene** Nicht-Terminalsymbole entweder für einen **Tokennamen** oder einen **Teil der Beschreibung** eines **Tokennamen**. Zum Beispiel handelt es sich bei dem **großgeschriebenen** Nicht-Terminalsymbol NUM um einen **Tokennamen**, der durch die **Produktion** `NUM ::= "0" | DIG.NO.0 DIG.WITH.0*` beschrieben wird und beschreibt, wie ein möglicher **Tokenwert**, in diesem Fall eine **Zahl** aufgebaut sein kann. Das ist daran festzumachen, dass das Nicht-Terminalsymbol NUM in keiner anderen Produktion vorkommt, die auf der **linken Seite** des „**kann abgeleitet werden zu**“-Symbols ::= ebenfalls ein **großgeschriebenen** Nicht-Terminalsymbol hat. Dagegen dient das **großgeschriebene** Nicht-Terminalsymbol DIG.NO.0 aus der Produktion `NUM ::= "0" | DIG.NO.0 DIG.WITH.0*` nur zu Beschreibung von NUM.

Die in der Grammatik 1.1.1 definierten **Nicht-Terminalsymbole** können in der Grammatik 1.2.8 der **Konkreten Syntax** für die **Syntaktischen Analyse** verwendet werden, um z.B. zu beschreiben, in welchem Kontext z.B. eine Zahl NUM stehen darf.

Die in der Konkrete Syntax vereinzelt **kleingeschriebenen** Nicht-Terminalsymbole, wie `name` haben nur den Zweck mehrere **Tokennamen**, wie `NAME | INT_NAME | CHAR_NAME` unter einem Überbegriff zu sammeln.

In Lark steht eine Zahl `.ZAHL`, die an ein **Nicht-Terminalsymbol** angehängt ist, dass auf der linken Seite des „**kann abgeleitet werden zu**“-Symbols ::= einer Produktion steht für die **Priorität** der Produktion dieses **Nicht-Terminalsymbols**. Es gibt den Fall, dass ein Wort von mehreren Produktionen erkannt wird, z.B. wird das Wort `int` sowohl von der Produktion `NAME`, als auch von der Produktion `INT.DT` erkannt. Daher ist es notwendig für `INT.DT` eine **Priorität** `INT.DT.2` zu setzen⁶, damit das Wort `int` den **Tokennamen** `INT.DT` zugewiesen bekommt und nicht `NAME`.

Allerdings muss für den Fall, dass `int` der **Präfix** eines Wortes ist, z.B. `int_var` noch die Produktion `INT.NAME.3` definiert werden, da der im **Lark Parsing Toolkit** verwendete **Basic Lexer** sobald ein Wort von einer Produktion erkannt wird, diesem direkt einen Tokennamen zuordnet, auch wenn das Wort eigentlich von einer anderen Produktion erkannt werden sollte. In diesem Fall würden aus `int_var` die Token `Token('INT_DT', 'int')`, `Token('NAME', '_var')` generiert, anstatt `Token(NAME, 'int_var')`. Daher muss die Produktion `INT.NAME.3` eingeführt werden, die immer zuerst geprüft wird. Wenn es sich nur um das Wort `int` handelt, wird zuerst die Produktion `INT.NAME.3` geprüft, es stellt sich heraus, dass `int` von der Produktion `INT.NAME.3` nicht erkannt wird, daher wird als nächstes `INT.DT.2` geprüft, welches `int` erkennt.

Der **Basic Lexer** des **Lark Parsing Toolkit** funktioniert grundlegend so wie es im Unterkapitel ?? erklärt wurde, allerdings berücksichtigt der **Basic Lexer** ebenfalls **Prioritäten**, sodass für den aktuellen Index im Eingabeprogramm zuerst alle Produktionen der **höchsten Priorität** geprüft werden. Sobald eine dieser Produktionen ein **Wort** an dem aktuellen Index im Eingabeprogramm erkennt, bekommt es direkt den

⁶Es wird immer die **höchste** Priorität **zuerst** genommen.

Tokenwert dieser Produktion zugewiesen, weitere Produktionen werden **nicht** mehr geprüft. Ansonsten werden alle Produktionen der **nächstniedrigeren** Priorität geprüft usw.

<i>COMMENT</i>	::=	"//"/[\wedge \backslash n]*/"/*"/(\cdot \backslash n)*?/"*/"	<i>L_Comment</i>
<i>RETI_COMMENT.2</i>	::=	"//"" \backslash "?"#"/[\wedge \backslash n]*/"	
<i>DIG_NO_0</i>	::=	"1" "2" "3" "4" "5" "6" "7" "8" "9"	<i>L_Arith</i>
<i>DIG_WITH_0</i>	::=	"0" <i>DIG_NO_0</i>	
<i>NUM</i>	::=	"0" <i>DIG_NO_0 DIG_WITH_0</i> *	
<i>ASCII_CHAR</i>	::=	" \backslash " \cdot " \backslash " \sim "	
<i>CHAR</i>	::=	"'" <i>ASCII_CHAR</i> ""	
<i>FILENAME</i>	::=	<i>ASCII_CHAR</i> + ". <i>picoc</i> "	
<i>LETTER</i>	::=	"a" \cdot "z" "A" \cdot "Z"	
<i>NAME</i>	::=	(<i>LETTER</i> " \backslash ") (<i>LETTER</i> <i>DIG_WITH_0</i> " \backslash ")*	
<i>name</i>	::=	<i>NAME</i> <i>INT_NAME</i> <i>CHAR_NAME</i> <i>VOID_NAME</i>	
<i>LOGIC_NOT</i>	::=	"!"	
<i>NOT</i>	::=	" \sim "	
<i>REF_AND</i>	::=	"&"	
<i>un_op</i>	::=	<i>SUB_MINUS</i> <i>LOGIC_NOT</i> <i>NOT</i> <i>MUL_DEREF_PNTR</i> <i>REF_AND</i>	
<i>MUL_DEREF_PNTR</i>	::=	"*"	
<i>DIV</i>	::=	"/"	
<i>MOD</i>	::=	"%"	
<i>prec1_op</i>	::=	<i>MUL_DEREF_PNTR</i> <i>DIV</i> <i>MOD</i>	
<i>ADD</i>	::=	"+"	
<i>SUB_MINUS</i>	::=	"-"	
<i>prec2_op</i>	::=	<i>ADD</i> <i>SUB_MINUS</i>	
<i>LT</i>	::=	"<"	<i>L_Logic</i>
<i>LTE</i>	::=	"<="	
<i>GT</i>	::=	">"	
<i>GTE</i>	::=	">="	
<i>rel_op</i>	::=	<i>LT</i> <i>LTE</i> <i>GT</i> <i>GTE</i>	
<i>EQ</i>	::=	"=="	
<i>NEQ</i>	::=	"!="	
<i>eq_op</i>	::=	<i>EQ</i> <i>NEQ</i>	
<i>INT_DT.2</i>	::=	" <i>int</i> "	<i>L_Assign_Alloc</i>
<i>INT_NAME.3</i>	::=	" <i>int</i> " (<i>LETTER</i> <i>DIG_WITH_0</i> " \backslash ")+	
<i>CHAR_DT.2</i>	::=	" <i>char</i> "	
<i>CHAR_NAME.3</i>	::=	" <i>char</i> " (<i>LETTER</i> <i>DIG_WITH_0</i> " \backslash ")+	
<i>VOID_DT.2</i>	::=	" <i>void</i> "	
<i>VOID_NAME.3</i>	::=	" <i>void</i> " (<i>LETTER</i> <i>DIG_WITH_0</i> " \backslash ")+	
<i>prim_dt</i>	::=	<i>INT_DT</i> <i>CHAR_DT</i> <i>VOID_DT</i>	

Grammatik 1.1.1: Grammatik der Konkreten Syntax der Sprache L_{Picoc} für die Lexikalische Analyse in EBNF

1.1.2 Codebeispiel

In den folgenden Unterkapiteln wird das Beispiel in Code 1.1 dazu verwendet die Konstruktion eines **Abstrakter Syntaxbaums** in seinen einzelnen **Zwischenschritten** zu erläutern.

```
1 struct st {int *(*attr)[4][5];};
2
3 void main() {
4     struct st *(*var[3][2]);
5 }
```

Code 1.1: PicoC-Code des Codebeispiels

Die vom **Basic Lexer** des **Lark Parsing Toolkit** erkannten **Token** sind Code 1.2 zu sehen.

```
1 [Token('FILENAME', './verbose_dt_simple_ast_gen_array_decl_and_alloc.pico'), Token('STRUCT',
  ↳ 'struct'), Token('NAME', 'st'), Token('LBRACE', '{'), Token('INT_DT', 'int'),
  ↳ Token('MUL_DEREF_PNTR', '*'), Token('LPAR', '('), Token('MUL_DEREF_PNTR', '*'),
  ↳ Token('NAME', 'attr'), Token('RPAR', ')'), Token('LSQB', '['), Token('NUM', '4'),
  ↳ Token('RSQB', ']'), Token('LSQB', '['), Token('NUM', '5'), Token('RSQB', ']'),
  ↳ Token('SEMICOLON', ';'), Token('RBRACE', '}'), Token('SEMICOLON', ';'), Token('VOID_DT',
  ↳ 'void'), Token('NAME', 'main'), Token('LPAR', '('), Token('RPAR', ')'), Token('LBRACE',
  ↳ '{'), Token('STRUCT', 'struct'), Token('NAME', 'st'), Token('MUL_DEREF_PNTR', '*'),
  ↳ Token('LPAR', '('), Token('MUL_DEREF_PNTR', '*'), Token('NAME', 'var'), Token('LSQB',
  ↳ '['), Token('NUM', '3'), Token('RSQB', ']'), Token('LSQB', '['), Token('NUM', '2'),
  ↳ Token('RSQB', ']'), Token('RPAR', ')'), Token('SEMICOLON', ';'), Token('RBRACE', '}')]
```

Code 1.2: Tokens für das Codebeispiel

1.2 Syntaktische Analyse

In der **Syntaktischen Analyse** ist es die Aufgabe des **Parsers** aus einem Programm in **Konkreter Syntax** unter Verwendung der **Tokens** aus der **Lexikalischen Analyse** einen **Ableitungsbaum** zu generieren. Es ist danach die Aufgabe möglicher **Visitors** und die Aufgabe des **Transformers** aus diesem **Ableitungsbaum** einen **Abstrakter Syntaxbaum** in **Abstrakter Syntax** zu generieren.

1.2.1 Umsetzung von Präzedenz und Assoziativität

Die Programmiersprache L_{PicoC} hat dieselben **Präzidenzregeln** implementiert, wie die Programmiersprache L_C ⁷. Die **Präzidenzregeln** der Programmiersprache L_{PicoC} sind in Tabelle 1.1 aufgelistet.

⁷C Operator Precedence - cppreference.com.

Präzidenzstufe	Operatoren	Beschreibung	Assoziativität
1	<code>a()</code>	Funktionsaufruf	Links, dann rechts \rightarrow
	<code>a[]</code>	Indezzugriff	
	<code>a.b</code>	Attributzugriff	
2	<code>-a</code>	Unäres Minus	Rechts, dann links \leftarrow
	<code>!a ~a</code>	Logisches NOT und Bitweise NOT	
	<code>*a &a</code>	Dereferenz und Referenz, auch Adresse-von	
3	<code>a*b a/b a%b</code>	Multiplikation, Division und Modulo	Links, dann rechts \rightarrow
4	<code>a+b a-b</code>	Addition und Subtraktion	
5	<code>a<b a<=b a>b a>=b</code>	Kleiner, Kleiner Gleich, Größer, Größer gleich	
6	<code>a==b a!=b</code>	Gleichheit und Ungleichheit	
7	<code>a&b</code>	Bitweise UND	
8	<code>a^b</code>	Bitweise XOR (exclusive or)	
9	<code>a b</code>	Bitweise ODER (inclusive or)	
10	<code>a&&b</code>	Logisches UND	
11	<code>a b</code>	Logisches ODER	Rechts, dann links \leftarrow
12	<code>a=b</code>	Zuweisung	

Tabelle 1.1: Präzidenzregeln von *PicoC*

Würde man diese **Operatoren** ohne Beachtung von **Präzidenzregeln** (Definition ??) und **Assoziativität** (Definition ??) in eine Grammatik verarbeiten wollen, so könnte eine Grammatik, wie Grammatik 1.2.1 dabei rauskommen.

<i>prim_exp</i>	::=	<i>name</i> <i>NUM</i> <i>CHAR</i> "(" <i>exp</i> ")"	<i>L_Arith</i> +
<i>un_op</i>	::=	"-" "~" "!" "*" "&"	
<i>un_exp</i>	::=	<i>un_op exp</i>	
<i>bin_op</i>	::=	"*" "/" "%" "+" "-" "&" "^" " " "<" "<=" ">" ">=" "!=" "==" "&&" " "	
<i>bin_exp</i>	::=	<i>exp bin_op exp</i>	
<i>exp</i>	::=	<i>prim_exp</i> <i>un_exp</i> <i>bin_exp</i>	

Grammatik 1.2.1: Undurchdachte Konkrete Syntax der Sprache L_{PicoC} für die Syntaktische Analyse in EBNF

Die Grammatik 1.2.1 ist allerdings **mehrdeutig**, d.h. verschiedene **Linksableitungen** in der Grammatik können zum selben **Wort** abgeleitet werden. Z.B. kann das Wort `3 * 1 && 4` sowohl über die **Linksableitung 1.2.1** als auch über die **Linksableitung 1.2.2** abgeleitet werden.

$$\begin{aligned}
 \text{exp} &\Rightarrow \text{bin_exp} \Rightarrow \text{exp bin_op exp} \Rightarrow \text{bin_exp bin_op exp} \\
 &\Rightarrow \text{exp bin_op exp bin_op exp} \Rightarrow^* 3 * 1 \&\& 4
 \end{aligned}
 \tag{1.2.1}$$

$$\begin{aligned}
 \text{exp} &\Rightarrow \text{bin_exp} \Rightarrow \text{exp bin_op exp} \Rightarrow \text{prim_exp bin_op exp} \Rightarrow \text{NUM bin_op exp} \\
 &\Rightarrow 3 \text{ bin_op exp} \Rightarrow 3 * \text{exp} \Rightarrow 3 * \text{bin_exp} \Rightarrow 3 * \text{exp bin_exp} \Rightarrow^* 3 * 1 \&\& 4
 \end{aligned}
 \tag{1.2.2}$$

Beide **Wörter** sind **gleich**, allerdings sind die **Ableitungsbäume unterschiedlich**, wie in Abbildung 1.1 zu sehen ist.

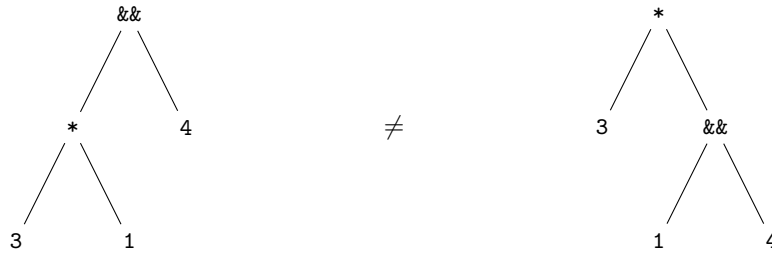


Abbildung 1.1: Ableitungsbäume zu den beiden Ableitungen

Der **linke Baum** entspricht Ableitung 1.2.1 und der **rechte Baum** entspricht Ableitung 1.2.2. Würde man in den Ausdrücken, die von diesen Bäumen dargestellt sind in **Klammern** setzen, um die **Präzidenz** sichtbar zu machen, so würde Ableitung 1.2.1 die Klammerung $(3 * 1) \&\& 4$ haben und die Ableitung 1.2.2 die Klammerung $3 * (1 \&\& 4)$ haben.

Aus diesem Grund ist es wichtig die **Präzidenzregeln** und die **Assoziativität** der Operatoren beim Erstellen der Grammatik miteinzubeziehen. Hierzu wird nun Tabelle 1.1 betrachtet. Für jede **Präzidenzstufe** in der Tabelle 1.1 wird eine eigene Regel erstellt werden, wie es in Grammatik 1.2.2 dargestellt ist. Zudem braucht es eine **Produktion** `prim.exp` für die höchste **Präzidenzstufe**, welche **Literale**, wie 'c', 5 oder `var` und geklammerte Ausdrücke wie $(3 \&\& 14)$ abdeckt.

<code>prim.exp</code>	::=	...	<code>L_Arith + L_Array</code>
<code>post.exp</code>	::=	...	<code>+ L_Pntr + L_Struct</code>
<code>un.exp</code>	::=	...	<code>+ L_Fun</code>
<code>arith.prec1</code>	::=	...	
<code>arith.prec2</code>	::=	...	
<code>arith.and</code>	::=	...	
<code>arith.oplus</code>	::=	...	
<code>arith.or</code>	::=	...	
<code>rel.exp</code>	::=	...	<code>L_Logic</code>
<code>eq.exp</code>	::=	...	
<code>logic.and</code>	::=	...	
<code>logic.or</code>	::=	...	
<code>assign.stmt</code>	::=	...	<code>L_Assign</code>

Grammatik 1.2.2: Durchdachte Konkrete Syntax der Sprache L_{PicoC} für die Syntaktische Analyse in EBNF

Einigen **Bezeichnungen** der **Produktionen** sind in Tabelle 1.2 ihren jeweiligen **Operatoren** zugeordnet für welche sie zuständig sind.

Bezeichnung der Produktionsregel	Operatoren
post_exp	a() a[] a.b
un_exp	-a !a ~a *a &a
arith_prec1	a*b a/b a%b
arith_prec2	a+b a-b
arith_and	a<b a<=b a>b a>=b
arith_oplus	a==b a!=b
arith_or	a&b
rel_exp	a^b
eq_exp	a b
logic_and	a&&b
logic_or	a b
assign	a=b

Tabelle 1.2: Zuordnung der Bezeichnungen von Produktionsregeln zu Operatoren

Als nächstes müssen die einzelnen **Produktionen** entsprechend der **Ausdrücke** für die sie zuständig sind definiert werden. Jede der **Produktionen** soll nur Ausdrücke **erkennen** können, deren **Präzidenzstufe** die ist, für welche die jeweilige Produktion verantwortlich ist oder deren Präzidenzstufe **höher** ist. Z.B. soll **un_op** sowohl den Ausdruck $-(3 * 14)$ als auch einfach nur $(3 * 14)$ ⁸ erkennen können, aber nicht $3 * 14$ ohne Klammern, da dieser Ausdruck eine **geringe Präzidenz** hat. Des Weiteren muss bei Produktionen für Ausdrücke mit **Operatoren** unterschieden werden, ob die Operatoren **linksassoziativ** oder **rechtsassoziativ**, **unär**, **binär** usw. sind.

Bei z.B. der Produktion **un_exp** in 1.2.3 für die **rechtsassoziativen unären Operatoren** $-a$, $!a$, $\sim a$, $*a$ und $\&a$ ist die **Alternative** **un_op un_exp** dafür zuständig, dass diese unären Operatoren **rechtsassoziativ** geschachtelt werden können (z.B. $! \sim 42$). Die Alternative **post_exp** ist dafür zuständig, dass die Produktion auch **terminieren** kann und es auch möglich ist ausschließlich einen Ausdruck **höherer Präzidenz** (z.B. 42) zu haben.

$$un_exp ::= un_op\ un_exp \mid post_exp$$

Grammatik 1.2.3: Beispiel für eine unäre rechtsassoziative Produktion

Bei z.B. der Produktion **post_exp** in 1.2.4 für die **linksassoziativen unären Operatoren** $a()$, $a[]$ und $a.b$ sind die Alternativen **post_exp**["logic_or"] und **post_exp**."name" dafür zuständig, dass diese unären Operatoren **linksassoziativ** geschachtelt werden können (z.B. $ar[3][1].car[4]$). Die Alternative **name**("fun_args") ist für einen **einzelnen Funktionsaufruf** zuständig. Die Alternative **prim_exp** ist dafür zuständig, dass die Produktion nicht nur bei **name**("fun_args") **terminieren** kann und es auch möglich ist ausschließlich einen Ausdruck der **höchsten Präzidenz** (z.B. 42) zu haben.

$$post_exp ::= post_exp\ ["logic_or"] \mid post_exp\ ."name" \mid name\ ("fun_args") \mid prim_exp$$

Grammatik 1.2.4: Beispiel für eine unäre linksassoziative Produktion

Bei z.B. der Produktion **prec2_exp** in 1.2.5 für die **binären linksassoziativen Operatoren** $a+b$ und $a-b$ ist die **Alternative** **arith_prec2 prec2_op arith_prec1** dafür zuständig, dass **mehrere** Operationen der Präzidenzstufe 4 in Folge erkannt werden können⁹ (z.B. $3 + 1 - 4$, wobei $-$ und $+$ beide Präzidenzstufe 4

⁸Geklammerte Ausdrücke werden nämlich von **prim_exp** erkannt, welches eine höhere **Präzidenzstufe** hat.

⁹Bezogen auf Tabelle 1.1.

haben). Das **Nicht-Terminalsymbol** `arith_prec1` auf der **rechten Seite** ermöglicht es, dass zwischen den Operationen der Präzidenzstufe 4 auch Operationen der Präzidenzstufe 3 auftauchen können (z.B. $3 + 1 / 4 - 1$, wobei $-$ und $+$ beide Präzidenzstufe 4 haben und $/$ Präzidenzstufe 3). Mit der Alternative `arith_prec1` ist es möglich, dass ausschließlich ein Ausdruck **höherer Präzidenz** erkannt wird (z.B. $1 / 4$).

$$\text{arith_prec2} ::= \text{arith_prec2 } \text{prec2_op } \text{arith_prec1} \mid \text{arith_prec1}$$

Grammatik 1.2.5: *Beispiel für eine linksassoziative Produktion*

Manche **Parser**^a haben allerdings ein Problem mit **Linksrekursion** (Definition ??), wie sie z.B. in der Produktion 1.2.5 vorliegt. Dieses Problem lässt sich allerdings einfach lösen, indem man die Produktion 1.2.5 zur Produktion 1.2.6 umschreibt.

$$\text{arith_prec2} ::= \text{arith_prec1 } (\text{prec2_op } \text{arith_prec1})^*$$

Grammatik 1.2.6: *Beispiel für eine linksassoziative Produktion*

Die von Produktion 1.2.6 erkannten Ausdrücke sind dieselben, wie für die Produktion 1.2.5, allerdings ist die Produktion 1.2.6 **flach** gehalten und ruft sich **nicht** selber auf, sondern nutzt den in der EBNF (Definition 1.3) definierten $*$ -Operator, um mehrere Operationen der Präzidenzstufe 4 in Folge erkennen zu können (z.B. $3 + 1 - 4$, wobei $-$ und $+$ beide Präzidenzstufe 4 haben).

Das **Nicht-Terminalsymbol** `arith_prec1` erlaubt es, dass **zwischen** der Folge von Operationen der Präzidenzstufe 4 auch Operationen der Präzidenzstufe 3 auftauchen können (z.B. $3 + 1 / 4 - 1$, wobei $-$ und $+$ beide Präzidenzstufe 4 haben und $/$ Präzidenzstufe 3). Da der in der EBNF definierte $*$ -Operator auch bedeutet, dass das Teilpattern auf das er sich bezieht **kein einziges mal** vorkommen kann, ist es mit dem **linken Nicht-Terminalsymbol** `arith_prec1` möglich, dass ausschließlich ein Ausdruck **höherer Präzidenz** erkannt wird (z.B. $1 / 4$).

^aDarunter zählt der **Earley Parser**, der im **PicoC-Compiler** verwendet wird **nicht**.

Alle **Operatoren** der Sprache L_{PicoC} sind also entweder **binär** und **linksassoziativ** (z.B. $a*b$, $a-b$, $a>b$ oder $a\&\&b$), **unär** und **rechtsassoziativ** (z.B. $\&a$ oder $!a$) oder **unär** und **linksassoziativ** (z.B. $a[]$ oder $a()$). Somit ergibt sich die Grammatik 1.2.7.

<i>prec1_op</i>	::=	"*" "/" "%"	<i>L_Misc</i>
<i>prec2_op</i>	::=	"+" "-"	
<i>rel_op</i>	::=	"<" "<=" ">" ">="	
<i>eq_op</i>	::=	"==" "!="	
<i>fun_args</i>	::=	[<i>logic_or</i> ("," <i>logic_or</i>)*]	
<i>prim_exp</i>	::=	<i>name</i> <i>NUM</i> <i>CHAR</i> "(" <i>logic_or</i> ")"	<i>L_Arith</i>
<i>post_exp</i>	::=	<i>post_exp</i> [" <i>logic_or</i> "] <i>post_exp</i> ." <i>name</i> " <i>name</i> (" <i>fun_args</i> ")"	+ <i>L_Array</i>
		<i>prim_exp</i>	+ <i>L_Pntr</i>
<i>un_exp</i>	::=	<i>un_op</i> <i>un_exp</i> <i>post_exp</i>	+ <i>L_Struct</i>
<i>arith_prec1</i>	::=	<i>arith_prec1</i> <i>prec1_op</i> <i>un_exp</i> <i>un_exp</i>	+ <i>L_Fun</i>
<i>arith_prec2</i>	::=	<i>arith_prec2</i> <i>prec2_op</i> <i>arith_prec1</i> <i>arith_prec1</i>	
<i>arith_and</i>	::=	<i>arith_and</i> "&" <i>arith_prec2</i> <i>arith_prec2</i>	
<i>arith_oplus</i>	::=	<i>arith_oplus</i> "^" <i>arith_and</i> <i>arith_and</i>	
<i>arith_or</i>	::=	<i>arith_or</i> " " <i>arith_oplus</i> <i>arith_oplus</i>	
<i>rel_exp</i>	::=	<i>rel_exp</i> <i>rel_op</i> <i>arith_or</i> <i>arith_or</i>	<i>L_Logic</i>
<i>eq_exp</i>	::=	<i>eq_exp</i> <i>eq_op</i> <i>rel_exp</i> <i>rel_exp</i>	
<i>logic_and</i>	::=	<i>logic_and</i> "&&" <i>eq_exp</i> <i>eq_exp</i>	
<i>logic_or</i>	::=	<i>logic_or</i> " " <i>logic_and</i> <i>logic_and</i>	
<i>assign_stmt</i>	::=	<i>un_exp</i> "=" <i>logic_or</i> ";"	<i>L_Assign</i>

Grammatik 1.2.7: Durchdachte Konkrete Syntax für Operatorpräzedenz in EBNF

1.2.2 Konkrete Syntax für die Syntaktische Analyse

Die gesamte Grammatik 1.2.8, welche die **Konkrete Syntax** der Sprache L_{PicoC} für die **Syntaktische Analyse** beschreibt ergibt sich wenn man die Grammatik 1.2.7 um die **restliche Syntax** der Sprache L_{PicoC} erweitert, die sich nach einem **ähnlichen Prinzip** wie in Unterkapitel 1.2.7 erläutert ergibt.

Später in der Entwicklung des **PicoC-Compilers** wurde die **Konkrete Syntax** an die **aktuellste konsistentlos auffindbare Version** der echten Grammatik *ANSI C grammar (Yacc)* der Sprache L_C angepasst¹⁰, damit es sicherer gewährleistet werden kann, dass der **PicoC-Compiler** sich genauso verhält, wie geläufige Compiler der Programmiersprache L_C , wobei z.B. die Compiler **GCC**¹¹ und **Clang**¹² zu nennen wären.

In der Grammatik 1.2.8, welche die **Konkrete Syntax** der Sprache L_{PicoC} für die **Syntaktische Analyse** beschreibt, werden einige der **Tokennamen** aus der Grammatik 1.1.1 der **Konkreten Syntax** für die **Lexikalischen Analyse** verwendet, wie z.B. *NUM* aber auch *name*, welches eine Produktion ist, die mehrere **Tokennamen** unter einem Überbegriff zusammenfasst.

Terminalsymbole, wie ; oder && gehören eigentlich zur **Lexikalischen Analyse**, jedoch erlaubt das **Lark Parsing Toolkit** um die Grammatik leichter lesbar zu machen einige **Terminalsymbole** einfach direkt in die Grammatik 1.2.8 der **Konkreten Syntax** für die **Syntaktische Analyse** zu schreiben. Der **Tokenname** für diese Terminalsymbole wird in diesem Fall vom **Lark Parsing Toolkit** bestimmt, welches einige sehr häufige verwendete **Terminalsymbole**, wie ; oder && bereits einen **Tokennamen** zugewiesen hat.

¹⁰An der für die Programmiersprache L_{PicoC} relevanten **Syntax** hat sich allerdings über die Jahre nichts verändert, wie die Grammatiken für die **Syntaktische Analyse** *ANSI C grammar (Lex)* und **Lexikalische Analyse** *noauthor'ansi'nodate-2* aus dem Jahre 1985 zeigen.

¹¹*GCC, the GNU Compiler Collection - GNU Project.*

¹²*clang: C++ Compiler.*

<i>prim_exp</i>	::=	<i>name</i> <i>NUM</i> <i>CHAR</i> "(" <i>logic_or</i> ")"	<i>L_Arith</i> + <i>L_Array</i>
<i>post_exp</i>	::=	<i>array_subscr</i> <i>struct_attr</i> <i>fun_call</i>	+ <i>L_Pntr</i> + <i>L_Struct</i>
		<i>input_exp</i> <i>print_exp</i> <i>prim_exp</i>	+ <i>L_Fun</i>
<i>un_exp</i>	::=	<i>un_op un_exp</i> <i>post_exp</i>	
<i>input_exp</i>	::=	"input" "(" ")"	
<i>print_exp</i>	::=	"print" "(" <i>logic_or</i> ")"	
<i>arith_prec1</i>	::=	<i>arith_prec1 prec1_op un_exp</i> <i>un_exp</i>	
<i>arith_prec2</i>	::=	<i>arith_prec2 prec2_op arith_prec1</i> <i>arith_prec1</i>	
<i>arith_and</i>	::=	<i>arith_and</i> "&" <i>arith_prec2</i> <i>arith_prec2</i>	
<i>arith_oplus</i>	::=	<i>arith_oplus</i> "^" <i>arith_and</i> <i>arith_and</i>	
<i>arith_or</i>	::=	<i>arith_or</i> " " <i>arith_oplus</i> <i>arith_oplus</i>	
<i>rel_exp</i>	::=	<i>rel_exp rel_op arith_or</i> <i>arith_or</i>	<i>L_Logic</i>
<i>eq_exp</i>	::=	<i>eq_exp eq_op rel_exp</i> <i>rel_exp</i>	
<i>logic_and</i>	::=	<i>logic_and</i> "&&" <i>eq_exp</i> <i>eq_exp</i>	
<i>logic_or</i>	::=	<i>logic_or</i> " " <i>logic_and</i> <i>logic_and</i>	
<i>type_spec</i>	::=	<i>prim_dt</i> <i>struct_spec</i>	<i>L_Assign_Alloc</i>
<i>alloc</i>	::=	<i>type_spec pntr_decl</i>	
<i>assign_stmt</i>	::=	<i>un_exp</i> "=" <i>logic_or</i> ";"	
<i>initializer</i>	::=	<i>logic_or</i> <i>array_init</i> <i>struct_init</i>	
<i>init_stmt</i>	::=	<i>alloc</i> "=" <i>initializer</i> ";"	
<i>const_init_stmt</i>	::=	"const" <i>type_spec name</i> "=" <i>NUM</i> ";"	
<i>pntr_deg</i>	::=	"*" *	<i>L_Pntr</i>
<i>pntr_decl</i>	::=	<i>pntr_deg array_decl</i> <i>array_decl</i>	
<i>array_dims</i>	::=	("[" <i>NUM</i> "]") *	<i>L_Array</i>
<i>array_decl</i>	::=	<i>name array_dims</i> "(" <i>pntr_decl</i> ")" <i>array_dims</i>	
<i>array_init</i>	::=	"{" <i>initializer</i> ("," <i>initializer</i>) * "}"	
<i>array_subscr</i>	::=	<i>post_exp</i> "[" <i>logic_or</i> "]"	
<i>struct_spec</i>	::=	"struct" <i>name</i>	<i>L_Struct</i>
<i>struct_params</i>	::=	(<i>alloc</i> ";") +	
<i>struct_decl</i>	::=	"struct" <i>name</i> "{" <i>struct_params</i> "}"	
<i>struct_init</i>	::=	"{" " ." <i>name</i> "=" <i>initializer</i> ("," " ." <i>name</i> "=" <i>initializer</i>) * "}"	
<i>struct_attr</i>	::=	<i>post_exp</i> " ." <i>name</i>	
<i>if_stmt</i>	::=	"if" "(" <i>logic_or</i> ")" <i>exec_part</i>	<i>L_If_Else</i>
<i>if_else_stmt</i>	::=	"if" "(" <i>logic_or</i> ")" <i>exec_part</i> "else" <i>exec_part</i>	
<i>while_stmt</i>	::=	"while" "(" <i>logic_or</i> ")" <i>exec_part</i>	<i>L_Loop</i>
<i>do_while_stmt</i>	::=	"do" <i>exec_part</i> "while" "(" <i>logic_or</i> ")" ";"	

Grammatik 1.2.8: Grammatik der Konkreten Syntax der Sprache L_{PicoC} für die Syntaktische Analyse in EBNF, Teil 1

<code>decl_exp_stmt</code>	<code>::= alloc";"</code>	<code>L_Stmt</code>
<code>decl_direct_stmt</code>	<code>::= assign_stmt init_stmt const_init_stmt</code>	
<code>decl_part</code>	<code>::= decl_exp_stmt decl_direct_stmt RETI_COMMENT</code>	
<code>compound_stmt</code>	<code>::= "{" exec_part * "}"</code>	
<code>exec_exp_stmt</code>	<code>::= logic_or";"</code>	
<code>exec_direct_stmt</code>	<code>::= if_stmt if_else_stmt while_stmt do_while_stmt</code>	
	<code> assign_stmt fun_return_stmt</code>	
<code>exec_part</code>	<code>::= compound_stmt exec_exp_stmt exec_direct_stmt</code>	
	<code> RETI_COMMENT</code>	
<code>decl_exec_stmts</code>	<code>::= decl_part * exec_part*</code>	
<code>fun_args</code>	<code>::= [logic_or(" ", logic_or)*]</code>	<code>L_Fun</code>
<code>fun_call</code>	<code>::= name("fun_args")</code>	
<code>fun_return_stmt</code>	<code>::= "return" [logic_or];"</code>	
<code>fun_params</code>	<code>::= [alloc(" ", alloc)*]</code>	
<code>fun_decl</code>	<code>::= type_spec ptr_deg name("fun_params")</code>	
<code>fun_def</code>	<code>::= type_spec ptr_deg name("fun_params") "{" decl_exec_stmts "}"</code>	
<code>decl_def</code>	<code>::= (struct_decl fun_decl);" fun_def</code>	<code>L_File</code>
<code>decls_defs</code>	<code>::= decl_def*</code>	
<code>file</code>	<code>::= FILENAME decls_defs</code>	

Grammatik 1.2.9: Grammatik der Konkreten Syntax der Sprache L_{PicoC} für die Syntaktische Analyse in EBNF, Teil 2

In der Grammatik 1.2.8 sind alle **Grammatiksymbole** ausgegraut, die das **Bachelorprojekt** betreffen. Alle nicht ausgegrauten **Grammatiksymbole** wurden für die Implementierung der **neuen Funktionalitäten**, welche die **Bachelorarbeit** betreffen hinzugefügt.

1.2.3 Ableitungsbaum Generierung

Die in Unterkapitel 1.2.2 definierte **Konkrete Syntax**, die von der Grammatik 1.2.8 beschrieben wird lässt sich mithilfe des **Earley Parsers** (Definition 1.6) von Lark dazu verwenden Code, der in der Sprache L_{PicoC} geschrieben ist zu parsen um einen **Ableitungsbaum** zu generieren.

Definition 1.6: Earley Parser

Ist ein Algorithmus für das **Parsen** von **Wörtern** einer **Kontextfreien Sprache**, der ein **Chart Parser** ist, welcher einen **top-down** arbeitenden **Earley Recognizer** (Defintion 1.7) nutzt, um einen **Ableitungsbaum** zu konstruieren.

Definition 1.7: Earley Recognizer

Ist ein **Recognizer**, der für **alle Kontextfreien Sprachen** das **Wortproblem** entscheiden kann und dies mittels **Dynamischer Programmierung** mit dem **Top-Down Ansatz** umsetzt.^a

Eingabe und **Ausgabe** des Algorithmus sind:

- **Eingabe:** Eingabewort w und **Grammatik** $G_{Parse} = \langle N, \Sigma, P, S \rangle$
- **Ausgabe:** 0 wenn $w \notin L(G_{Parse})^b$ und 1 wenn $w \in L(G_{Parse})$

Bevor dieser **Algorithmus** erklärt wird müssen noch einige **Symbole** und **Notationen** erklärt werden:

- α, β, γ stellen eine **beliebige Folge** von **Grammatiksymbolen**^c dar
- A und B stellen **Nicht-Terminalsymbole** dar
- a stellt ein **Terminalsymbol** dar
- **Earley's Punktnotation:** $A ::= \alpha \bullet \beta$ stellt eine **Produktion**, in der α **bereits geparkt** wurde und β **noch geparkt** werden muss
- Die **Indexierung** ist informell ausgedrückt so umgesetzt, dass die Indices **zwischen Tokennamen** liegen, also **Index 0** vor dem ersten **Tokennamen** verortet ist, **Index 1** nach dem ersten **Tokennamen** verortet ist und **Index n** nach dem **letzten Tokennamen** verortet ist

und davor müssen noch einige **Begriffe** definiert werden:

- **Zustandsmenge:** Für jeden der $n + 1$ **Indices** j wird eine **Zustandsmenge** $Z(j)$ generiert
- **Zustand einer Zustandsmenge:** Ist ein **Tupel** $(A ::= \alpha \bullet \beta, i)$, wobei $A ::= \alpha \bullet \beta$ die **momentane Produktion** ist, die bis **Punkt \bullet** geparkt wurde und i der **Index** ist, ab welchem der Versuch der Erkennung eines **Teilworts** des **Eingabeworts** mithilfe dieser **Produktion** begann

Der **Ablauf** des Algorithmus ist wie folgt:

1. **initialisiere** $Z(0)$ mit der **Produktion**, welches das **Startsymbol** S auf der **linken Seite** des „kann abgeleitet werden zu“-Symbols $::=$ hat
2. es werden in der **aktuellen Zustandsmenge** $Z(j)$ die folgenden **Operationen** ausgeführt:
 - **Voraussage:** Für jeden **Zustand** in der **Zustandsmenge** $Z(j)$, der die Form $(A ::= \alpha \bullet B\gamma, i)$ hat, wird für jede **Produktion** $(B ::= \beta)$ in der Grammatik, die ein B auf der **linken Seite** des „kann abgeleitet werden zu“-Symbols $::=$ hat ein **Zustand** $(B ::= \bullet\beta, j)$ zur **Zustandsmenge** $Z(j)$ hinzugefügt
 - **Überprüfung:** Für jeden **Zustand** in der **Zustandsmenge** $Z(j)$, der die Form $(A ::= \alpha \bullet a\gamma, i)$ hat wird der **Zustand** $(A ::= \alpha a \bullet \gamma, i)$ zur **Zustandsmenge** $Z(j + 1)$ hinzugefügt
 - **Vervollständigung:** Für jeden **Zustand** in der **Zustandsmenge** $Z(j)$, der die Form $(B ::= \beta \bullet, i)$ hat werden alle **Zustände** in $Z(i)$ gesucht, welche die Form $(A ::= \alpha \bullet B\gamma, i)$ haben und es wird der **Zustand** $(A ::= \alpha B \bullet \gamma, i)$ zur **Zustandsmenge** $Z(j)$ hinzugefügt

bis:

- der **Zustand** $(A ::= \beta \bullet, 0)$ in der **Zustandsmenge** $Z(n)$ auftaucht, wobei A das **Startsymbol** S ist $\Rightarrow w \in L(G_{Parse})$
- **keine Zustände** mehr einer **Zustandsmenge** hinzugefügt werden können $\Rightarrow w \notin L(G_{Parse})$

^aEarley, „An efficient context-free parsing“.

^b $L(G_{Parse})$ ist die **Sprache**, welche durch die **Grammatik** G_{Parse} beschrieben wird.

^cAlso eine Folge von **Terminalsymbolen** und **Nicht-Terminalsymbolen**.

1.2.3.1 Codebeispiel

Der **Ableitungsbaum**, der mithilfe des **Earley Parsers** und der **Token** der **Lexikalischen Analyse** aus dem Beispiel in Code 1.1 generiert wurde, ist in Code 1.3 zu sehen. Im Code 1.3 wurden einige Zeilen **markiert**, die später in Unterkapitel 1.2.4.1 zum Vergleich wichtig sind.

```

1 file
2 ./verbose_dt_simple_ast_gen_array_decl_and_alloc.dt
3 decls_defs
4 decl_def
5 struct_decl
6     name      st
7     struct_params
8     alloc
9         type_spec
10        prim_dt      int
11        ptr_decl
12        ptr_deg      *
13        array_decl
14        ptr_decl
15        ptr_deg      *
16        array_decl
17        name      attr
18        array_dims
19        array_dims
20        4
21        5
22 decl_def
23 fun_def
24     type_spec
25     prim_dt      void
26     ptr_deg
27     name      main
28     fun_params
29     decl_exec_stmts
30     decl_part
31     decl_exp_stmt
32     alloc
33         type_spec
34         struct_spec
35         name      st
36         ptr_decl
37         ptr_deg      *
38         array_decl
39         ptr_decl
40         ptr_deg      *
41         array_decl
42         name      var
43         array_dims
44         3
45         2
46         array_dims

```

Code 1.3: Ableitungsbaum nach Ableitungsbaum Generierung

1.2.3.2 Ausgabe des Ableitungsbaums

Die Ausgabe des **Ableitungsbaums** wird komplett vom **Lark Parsing Toolkit** übernommen. Für die **Inneren Knoten** werden die **Nicht-Terminalsymbole**, welche in der Grammatik den **linken Seiten** des „**kann abgeleitet werden zu**“-**Symbols** $::=$ ¹³ entsprechen hergenommen und die **Blätter** sind **Terminalsymbole**, genauso, wie es in der Definition ?? eines **Ableitungsbaums** auch schon definiert ist. Die **EBNF-Grammatik 1.2.8** des **PicoC-Compilers** erlaubt es allerdings auch, dass in einem **Blatt** garnichts ε steht, weil es z.B. **Produktionen**, wie `array_dims ::= ("NUM")*` gibt, in denen auch das **leere Wort** ε abgeleitet werden kann.

Die Ausgabe des **Abstrakter Syntaxbaum** ist bewusst so gewählt, dass sie sich optisch vom **Ableitungsbaum** unterscheidet, indem die Bezeichner der **Knoten** in **UpperCamelCase** geschrieben sind, im Gegensatz zum **Ableitungsbaum**, dessen **Innere Knoten** im **snake_case** geschrieben sind, wie auch die **Nicht-Terminalsymbole** auf den **linken Seiten** des „**kann abgeleitet werden zu**“-**Symbols** $::=$.

1.2.4 Ableitungsbaum Vereinfachung

Der **Ableitungsbaum** in Code 1.3, dessen Generierung in Unterkapitel 1.2.3.1 besprochen wurde ist noch untauglich, damit aus ihm mittels eines **Transformers** ein **Abstrakter Syntaxbaum** generiert werden kann. Das Problem ist, dass um den den **Datentyp** einer Variable in der Programmiersprache L_C und somit auch die Programmiersprache L_{PicoC} korrekt bestimmen zu können, wie z.B. ein „**Array der Mächtigkeit 3 von Pointern auf Arrays der Mächtigkeit 2 von Integern**“ `int (*ar[3])[2]` die **Spiralregel**¹⁴ in der Implementierung des **PicoC-Compilers** umgesetzt werden muss und das ist nicht alleinig möglich, indem man die entsprechenden **Produktionen** in der Grammatik 1.2.8 der **Konkreten Syntax** auf eine spezielle Weise passend spezifiziert.

Was man erhalten will, ist ein **entarteter Baum** von **PicoC-Knoten**, an dem man den **Datentyp** direkt ablesen kann, indem man sich einfach über den **entarteten Baum** bewegt, wie z.B. `PntrDecl(Num('1'),ArrayDecl([Num('3'),Num('2')],PntrDecl(Num('1'),StructSpec(Name('st')))))` für den Ausdruck `struct st *(*var[3][2])`.

Es sind hierbei mehrere Probleme zu lösen. Hat man den Ausdruck `struct st *(*var[3][2])` wird dieser zu einem **Ableitungsbaum**, wie er in Abbildung 1.2 zu sehen ist.

¹³Grammar: The language of languages (BNF, EBNF, ABNF and more).

¹⁴Clockwise/Spiral Rule.

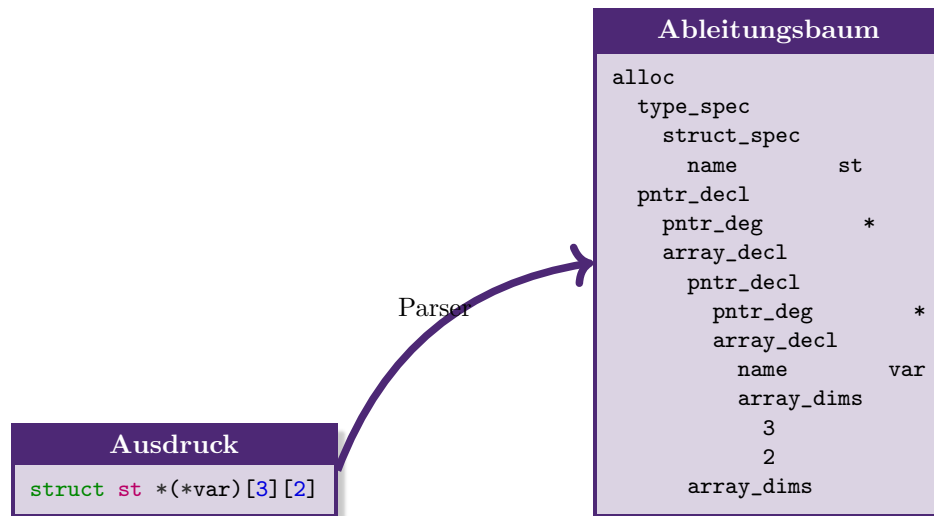


Abbildung 1.2: Ableitungsbaum nach Parsen eines Ausdrucks

Dieser **Ableitungsbaum** für den Ausdruck `struct st *(*var)[3][2]` hat allerdings einen Aufbau welcher durch die **Syntax** der **Pointerdeklaratoren** `ptr_decl(num, datatype)` und **Arraydeklaratoren** `array_decl(datatype, nums)` bestimmt ist, die **spiralähnlich** ist. Man würde allerdings gerne einen **entarteten Baum** erhalten, bei dem der Datentyp immer im **zweiten Attribut** weitergeht, anstatt abwechselnd im **zweiten** und **ersten**, wie beim **Pointerdeklarator** `ptr_decl(num, datatype)` und **Arraydeklarator** `array_decl(datatype, nums)`. Daher muss beim **ArrayDeclarator** `array_decl(datatype, nums)` immer das **erste Attribut** `datatype` mit dem **zweiten Attribut** `nums` getauscht werden.

Des Weiteren befindet sich in der **Mitte** dieser **Spirale**, die der **Ableitungsbaum** bildet der **Name der Variable** `name(var)` und nicht der **innerste Datentyp** `struct st`, da der **Ableitungsbaum** einfach nur die **kompilerinterne Darstellung**, die durch das Parsen eines **Programms in Konkreter Syntax** (z.B. `struct st *(*var)[3][2]`) **generiert** wird darstellt. Der **Name der Variable** `name(var)` sollte daher mit dem **innersten Datentyp** `struct st` ausgetauscht werden.

In Abbildung 1.3 ist daher zu sehen, wie der **Ableitungsbaum** aus Abbildung 1.2 mithilfe eines **Visitors** (Definition ??) **vereinfacht** wird, sodass er die gerade erläuterten Ansprüche erfüllt.

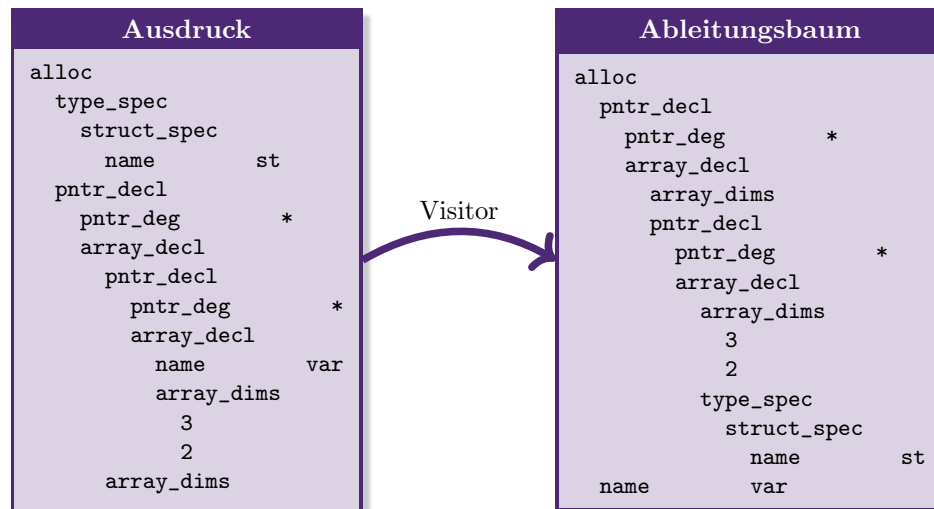


Abbildung 1.3: Ableitungsbaum nach Vereinfachung

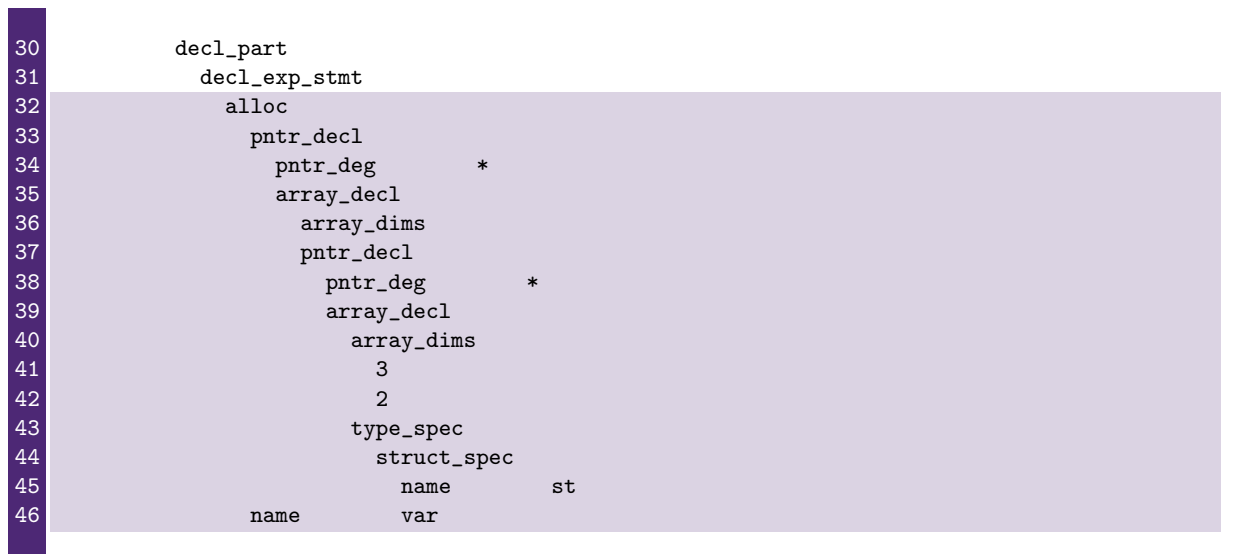
1.2.4.1 Codebeispiel

In Code 1.4 ist der **Ableitungsbaum** aus Code 1.3 nach der **Vereinfachung** mithilfe eines **Visitors** zu sehen.

```

1 file
2 ./verbose_dt_simple_ast_gen_array_decl_and_alloc.dt_simple
3 decls_defs
4   decl_def
5     struct_decl
6       name      st
7       struct_params
8         alloc
9           pntr_decl
10            pntr_deg      *
11            array_decl
12            array_dims
13            4
14            5
15            pntr_decl
16              pntr_deg      *
17              array_decl
18              array_dims
19              type_spec
20                prim_dt      int
21              name      attr
22      decl_def
23        fun_def
24          type_spec
25            prim_dt      void
26          pntr_deg
27            name      main
28          fun_params
29          decl_exec_stmts

```



Code 1.4: Ableitungsbaum nach Ableitungsbaum Vereinfachung

1.2.5 Abstrakt Syntax Tree Generierung

Nachdem der **Derivation Tree** in Unterkapitel 1.2.4 vereinfacht wurde, ist der **vereinfachte Ableitungsbaum** in Code 1.4 nun dazu geeignet, um mit einem **Transformer** (Definition ??) einen **Abstrakter Syntaxbaum** aus ihm zu generieren. Würde man den **vereinfachten Ableitungsbaum** des Ausdrucks `struct st *(*var[3][2])` auf passende Weise in einen **Abstrakter Syntaxbaum** umwandeln, so würde dabei ein **Abstrakter Syntaxbaum** wie in Abbildung 1.4 rauskommen.

Den Teilbaum, der den Datentyp darstellt würde man von von **oben-nach-unten**¹⁵ als „**Pointer auf einen Pointer auf ein Array der Mächtigkeit 2, 3 von Structs des Typs st**“ lesen, also genau anders herum, als man den Ausdruck `struct st *(*var[3][2])` mit der **Spiralregel** lesen würde. Bei der **Spiralregel** fängt man beim Ausdruck `struct st *(*var[3][2])` bei der Variable `var` an und arbeitet sich dann auf „**Spiralbahnen**“, von **innen-nach-außen** durch den Ausdruck, um herauszufinden, dass dieser Datentyp ein „**Array der Mächtigkeit 3, 2 von Pointern auf einen Pointer auf einen Struct vom Typ st**“ ist.

¹⁵In der Informatik wachsen Bäume von **oben-nach-unten**, von der **Wurzel** zur den **Blättern**, bzw. in diesem Beispiel von **links-nach-rechts**.

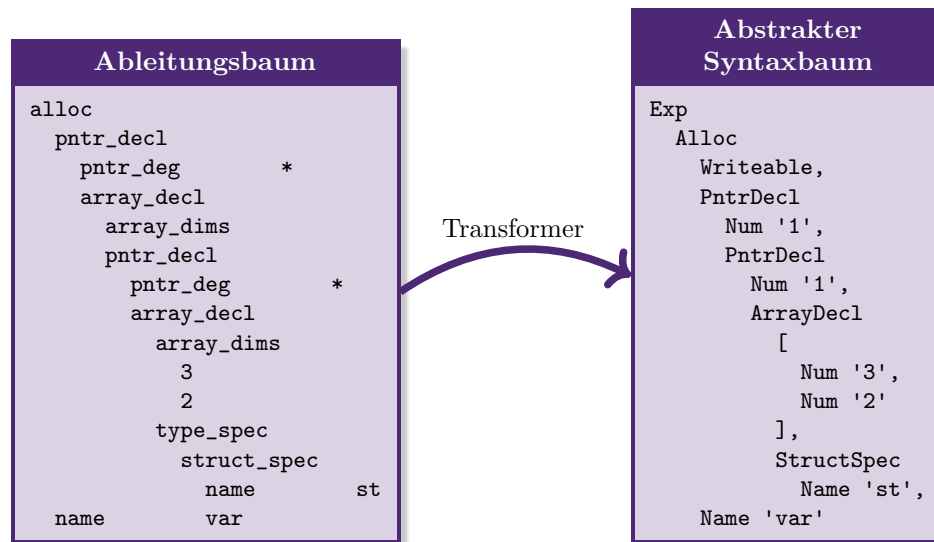


Abbildung 1.4: Abstrakter Syntaxbaum Generierung ohne Umdrehen

Dieser **Abstrakter Syntaxbaum** ist für die Weiterverarbeitung ungeeignet, denn für die **Adressberechnung** für eine Aneinanderreihung von Zugriffen auf **Pointerelemente**, **Arrayelemente** oder **Structattribute**, welche in Unterkapitel ?? genauer erläutert wird, will man den Datentyp in **umgekehrter Reihenfolge**. Aus diesem Grund muss der **Transformer** bei der Konstruktion des **Abstrakter Syntaxbaum** zusätzlich dafür sorgen, dass jeder **Teilbaum**, der für einen **Datentyp** steht **umgedreht** wird. Auf diese Weise kommt ein **Abstrakter Syntaxbaum** mit **richtig rum gedrehtem Datentyp**, wie in Abbildung 1.5 zustande, der für die Weiterverarbeitung geeignet ist.

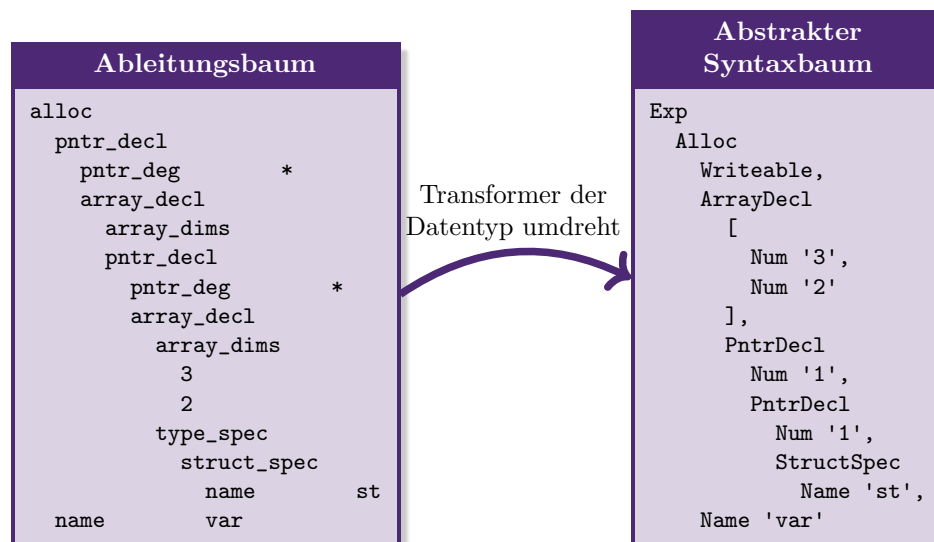


Abbildung 1.5: Abstrakter Syntaxbaum Generierung mit Umdrehen

Die Weiterverarbeitung des **Abstrakter Syntaxbaums** geschieht mithilfe von **Passes**, welche im Unterkapitel ?? genauer beschrieben werden. Da die Knoten des **Abstrakter Syntaxbaum** anders als beim **Ableitungsbaum** nicht die gleichen Bezeichnungen haben wie **Produktionen** der Grammatik der **Kon-**

kretten Syntax ist es in den folgenden Unterkapiteln 1.2.5.1, 1.2.5.2 und 1.2.5.3 notwendig die **Bedeutung** der einzelnen **PicoC-Knoten**, **RETI-Knoten** und bestimmter **Kompositionen** dieser Knoten zu **dokumentieren**, die alle in den unterschiedlichen von den **Passes** umgeformten **Abstrakter Syntaxbaums** vorkommen.

Des Weiteren gibt die **Abstrakte Syntax** die durch die Grammatik 1.2.1 in Unterkapitel 1.2.5.4 beschrieben wird aufschluss darüber welche **Kompositionen von PicoC-Knoten**, neben den bereits in Tabelle 1.2.10 definierten Kompositionen mit Bedeutung insgesamt überhaupt **möglich** sind.

1.2.5.1 PicoC-Knoten

Bei den **PicoC-Knoten** handelt es sich um Knoten, die irgendeinen **Ausdruck** aus der Sprache L_{PicoC} darstellen. Für die **PicoC-Knoten** wurden möglichst **kurze** und **leicht** verständliche Bezeichner gewählt, da auf diese Weise bei der Implementierung der einzelnen Passes möglichst **viel Code in eine Zeile** passt und dieser Code auch durch leicht verständliche Bezeichner von Knoten **intuitiv verständlich** sein sollte¹⁶. Alle **PicoC-Knoten**, die in den von den verschiedenen Passes generierten **Abstrakter Syntaxbaums** vorkommen sind in Tabelle 1.3 mit einem **Bschreibungstext** dokumentiert.

¹⁶Z.B. steht der **PicoC-Knoten** `Name(str)` für einen **Bezeichner**. Anstatt diesen Knoten in englisch `Identifizier(str)` zu nennen, wurde dieser als `Name(str)` gewählt, da `Name(str)` **kürzer** ist und **intuitiver verständlich**.

Pioco-Knoten	Beschreibung
Name(val)	Ein Bezeichner , z.B. <code>my_fun</code> , <code>my_var</code> usw., aber da es keine gute Kurzform für <code>Identifizier()</code> (englisches Wort für Bezeichner) gibt, wurde dieser Knoten <code>Name()</code> genannt.
Num(val)	Eine Zahl , z.B. 42, -3 usw.
Char(val)	Ein Zeichen der ASCII-Zeichenkodierung , z.B. <code>'c'</code> , <code>'*'</code> usw.
Minus(), Not(), DerefOp(), RefOp(), LogicNot()	Die unären Operatoren <code>un_op</code> : <code>-a</code> , <code>~a</code> , <code>*a</code> , <code>&a !a</code> .
Add(), Sub(), Mul(), Div(), Mod(), Oplus(), And(), Or(), LogicAnd(), LogicOr()	Die binären Operatoren <code>bin_op</code> : <code>a + b</code> , <code>a - b</code> , <code>a * b</code> , <code>a / b</code> , <code>a % b</code> , <code>a ^ b</code> , <code>a & b</code> , <code>a b</code> , <code>a && b</code> , <code>a b</code> .
Eq(), NEq(), Lt(), LtE(), Gt(), GtE()	Die Relationen <code>rel</code> : <code>a == b</code> , <code>a != b</code> , <code>a < b</code> , <code>a <= b</code> , <code>a > b</code> , <code>a >= b</code> .
Const(), Writeable()	Die Type Qualifier <code>type_qual</code> : <code>const</code> , was für ein nicht beschreibbare Konstante steht und das nicht Angeben von <code>const</code> , was für einen beschreibbare Variable steht.
IntType(), CharType(), VoidType()	Die Type Specifier für Primitiven Datentypen , die in der Abstrakten Syntax, um eine intuitive Bezeichnung zu haben einfach nur unter Datentypen <code>datatype</code> eingeordnet werden: <code>int</code> , <code>char</code> , <code>void</code> .
Placeholder()	Platzhalter für einen Knoten, der diesen später ersetzt .
BinOp(exp, bin_op, exp)	Container für eine binäre Operation mit 2 Expressions: <code><exp1> <bin_op> <exp2></code>
UnOp(un_op, exp)	Container für eine unäre Operation mit einer Expression: <code><un_op> <exp></code> .
Exit(num)	Container für einen Exit Code , der vor der Beendigung in das ACC Register geschrieben wird und steht für die Beendigung des laufenden Programmes.
Atom(exp, rel, exp)	Container für eine binäre Relation mit 2 Expressions: <code><exp1> <rel> <exp2></code>
ToBool(exp)	Container für einen Arithmetischen Ausdruck , wie z.B. <code>1 + 3</code> oder einfach nur <code>3</code> , der nicht nur 1 oder 0 als Ergebnis haben kann und daher bei einem Ergebnis <code>x > 1</code> auf 1 abgebildet wird.
Alloc(type_qual, datatype, name, local_var_or_param)	Container für eine Allokation <code><type_qual> <datatype> <name></code> mit den notwendigen Knoten <code>type_qual</code> , <code>datatype</code> und <code>name</code> , die alle für einen Eintrag in der Symboltabelle notwendigen Informationen enthalten. Zudem besitzt er ein verstecktes Attribut <code>local_var_or_param</code> , dass die Information trägt, ob es sich bei der Variable um eine Lokale Variable oder einen Parameter handelt.
Assign(lhs, exp)	Container für eine Zuweisung , wobei <code>lhs</code> ein <code>Subscr(exp1, exp2)</code> , <code>Deref(exp1, exp2)</code> , <code>Attr(exp, name)</code> oder <code>Name('var')</code> sein kann und <code>exp</code> ein beliebiger Logischer Ausdruck sein kann: <code>lhs = exp</code> .

Tabelle 1.3: PicoC-Knoten Teil 1

Piocc-Knoten	Beschreibung
Exp(exp, datatype, error_data)	Container für einen beliebigen Ausdruck , dessen Ergebnis auf den Stack soll. Zudem besitzt er 2 versteckte Attribute, wobei datatype im RETI Blocks Pass wichtig ist und error_data für Fehlermeldungen wichtig ist.
Stack(num)	Container, der für das temporäre Ergebnis einer Berechnung, das num Speicherzellen relativ zum Stackpointer Register SP steht.
Stackframe(num)	Container, der für eine Variable steht, die num Speicherzellen relativ zum Begin-Aktive-Funktion Register BAF steht.
Global(num)	Container, der für eine Variable steht, die num Speicherzellen relativ zum Datensegment Register DS steht.
StackMalloc(num)	Container, der für das Allokieren von num Speicherzellen auf dem Stack steht.
PntrDecl(num, datatype)	Container, der für den Pointerdatatype steht: <code><prim_dt>*<var></code> , wobei das Attribut num die Anzahl zusammengefasster Pointer angibt und datatype der Datentyp ist, auf den der oder die Pointer zeigen.
Ref(exp, datatype, error_data)	Container, der für die Anwendung des Referenz-Operators <code>&<var></code> steht und die Adresse einer Location (Definition ??) auf den Stack schreiben soll, die über exp eingegrenzt wird. Zudem besitzt er 2 versteckte Attribute, wobei datatype im RETI Blocks Pass wichtig ist und error_data für Fehlermeldungen wichtig ist.
Deref(lhs, exp)	Container für den Indexzugriff auf einen Array- oder Pointerdatatype : <code><var>[<i>]</code> , wobei exp1 eine angehängte weitere <code>Subscr(exp1, exp2)</code> , <code>Deref(exp1, exp2)</code> , <code>Attr(exp, name)</code> oder ein <code>Name('var')</code> sein kann und exp2 der Index ist auf den zugegriffen werden soll.
ArrayDecl(nums, datatype)	Container, der für den Arraydatatype steht: <code><prim_dt><var>[<i>]</code> , wobei das Attribut nums eine Liste von <code>Num('x')</code> ist, die die Dimensionen des Arrays angibt und datatype der Datentyp ist, der über das Anwenden von <code>Subscript()</code> auf das Array zugreifbar ist.
Array(exps, datatype)	Container für den Initializer eines Arrays , dessen Einträge exps weitere Initializer für eine Array-Dimension oder ein Initializer für ein Struct oder ein Logischer Ausdruck sein können, z.B. <code>{{1, 2}, {3, 4}}</code> . Des Weiteren besitzt er ein verstecktes Attribut datatype , welches für den PicoC-ANF Pass Informationen transportiert, die für Fehlermeldungen wichtig sind.
Subscr(exp1, exp2)	Container für den Indexzugriff auf einen Array- oder Pointerdatatype : <code><var>[<i>]</code> , wobei exp1 eine angehängte weitere <code>Subscr(exp1, exp2)</code> , <code>Deref(exp1, exp2)</code> oder <code>Attr(exp, name)</code> Operation sein kann oder ein <code>Name('var')</code> sein kann und exp2 der Index ist auf den zugegriffen werden soll.
StructSpec(name)	Container für einen selbst definierten Structdatatype : <code>struct <name></code> , wobei das Attribut name festlegt, welchen selbst definierte Structdatatype dieser Container-Knoten repräsentiert.
Attr(exp, name)	Container für den Attributzugriff auf einen Structdatatype : <code><var>.<attr></code> , wobei exp1 eine angehängte weitere <code>Subscr(exp1, exp2)</code> , <code>Deref(exp1, exp2)</code> oder <code>Attr(exp, name)</code> Operation sein kann oder ein <code>Name('var')</code> sein kann und name das Attribut ist, auf das zugegriffen werden soll.

Tabelle 1.4: PicoC-Knoten Teil 2

PiocC-Knoten	Beschreibung
Struct(assigns, datatype)	Container für den Initializer eines Structs , z.B. {.<attr1>={1, 2}, .<attr2>={3, 4}}, dessen Eintrag assigns eine Liste von Assign(lhs, exp) ist mit einer Zuordnung eines Attributezeichners , zu einem weiteren Initializer für eine Array-Dimension oder zu einem Initializer für ein Struct oder zu einem Logischen Ausdruck . Des Weiteren besitzt er ein verstecktes Attribut datatype, welches für den PicoC-ANF Pass Informationen transportiert, die für Fehlermeldungen wichtig sind.
StructDecl(name, allocs)	Container für die Deklaration eines selbstdefinierten Structdatentyps , z.B. struct <var> {<datatype> <attr1>; <datatype> <attr2>;}, wobei name der Bezeichner des Structdatentyps ist und allocs eine Liste von Bezeichnern der Attribute des Structdatentyps mit dazugehörigem Datentyp , wofür sich der Container-Knoten Alloc(type_qual, datatype, name) sehr gut als Container eignet.
If(exp, stmts)	Container für ein If Statement if(<exp>) { <stmts> } inklusive Condition exp und einem Branch stmts, indem eine Liste von Statements stehen kann oder ein einzelnes GoTo(Name('block.xyz')).
IfElse(exp, stmts1, stmts2)	Container für ein If-Else Statement if(<exp>) { <stmts2> } else { <stmts2> } inklusive Condition exp und 2 Branches stmts1 und stmts2, die zwei Alternativen darstellen in denen jeweils Listen von Statements oder GoTo(Name('block.xyz'))'s stehen können.
While(exp, stmts)	Container für ein While-Statement while(<exp>) { <stmts> } inklusive Condition exp und einem Branch stmts, indem eine Liste von Statements stehen kann oder ein einzelnes GoTo(Name('block.xyz')).
DoWhile(exp, stmts)	Container für ein Do-While-Statement do { <stmts> } while(<exp>); inklusive Condition exp und einem Branch stmts, indem eine Liste von Statements stehen kann oder ein einzelnes GoTo(Name('block.xyz')).
Call(name, exps)	Container für einen Funktionsaufruf : fun_name(exps), wobei name der Bezeichner der Funktion ist, die aufgerufen werden soll und exps eine Liste von Argumenten ist, die an die Funktion übergeben werden soll.
Return(exp)	Container für ein Return-Statement : return <exp>, wobei das Attribut exp einen Logischen Ausdruck darstellt, dessen Ergebnis vom Return-Statement zurückgegeben wird.
FunDecl(datatype, name, allocs)	Container für eine Funktionsdeklaration , z.B. <datatype> <fun_name>(<datatype> <param1>, <datatype> <param2>), wobei datatype der Rückgabewert der Funktion ist, name der Bezeichner der Funktion ist und allocs die Parameter der Funktion sind, wobei der Container-Knoten Alloc(type_spec, datatype, name) als Container für die Parameter dient.

Tabelle 1.5: PicoC-Knoten Teil 3

PiocC-Knoten	Beschreibung
FunDef(datatype, name, allocs, stmts.blocks)	Container für eine Funktionsdefinition , z.B. <datatype> <fun_name>(<datatype> <param>) {<stmts>}, wobei datatype der Rückgabewert der Funktion ist, name der Bezeichner der Funktion ist, allocs die Parameter der Funktion sind, wobei der Container-Knoten Alloc(type_spec, datatype, name) als Container für die Parameter dient und stmts.blocks eine Liste von Statements bzw. Blöcken ist, welche diese Funktion beinhaltet.
NewStackframe(fun_name, goto_after_call)	Container für die Erstellung eines neuen Stackframes und Speicherung des Werts des BAF-Registers der aufgerufenen Funktion und der Rücksprungadresse nacheinander an den Anfang des neuen Stackframes . Das Attribut fun_name steht dabei für den Bezeichner der Funktion, für die ein neuer Stackframe erstellt werden soll. Das Attribut fun_name dient später dazu den Block dieser Funktion zu finden, weil dieser für den weiteren Kompilierungsvorgang wichtige Information in seinen versteckten Attributen gespeichert hat. Des Weiteren ist das Attribut goto_after_call ein GoTo(Name('addr@next_instr')), welches später durch die Adresse des Befehls, der direkt auf die Jump Instruction folgt, ersetzt wird.
RemoveStackframe()	Container für das Entfernen des aktuellen Stackframes durch das Wiederherstellen des im noch aktuellen Stackframe gespeicherten Werts des BAF-Registers der aufgerufenen Funktion und das Setzen des SP-Registers auf den Wert des BAF-Registers vor der Wiederherstellung.
File(name, decls_defs.blocks)	Container für alle Funktionen oder Blöcke , welche eine Datei als Ursprung haben, wobei name der Dateiname der Datei ist, die erstellt wird und decls_defs.blocks eine Liste von Funktionen bzw. Blöcken ist.
Block(name, stmts_instrs, instrs_before, num_instrs, param_size, local_vars_size)	Container für Statements , der auch als Block bezeichnet wird, wobei das Attribut name der Bezeichner des Labels (Definition 1.8) des Blocks ist und stmts_instrs eine Liste von Statements oder Instructions . Zudem besitzt er noch 3 versteckte Attribute, wobei instrs_before die Zahl der Instructions vor diesem Block zählt, num_instrs die Zahl der Instructions ohne Kommentare in diesem Block zählt, param_size die voraussichtliche Anzahl an Speicherzellen aufaddiert, die für die Parameter der Funktion belegt werden müssen und local_vars_size die voraussichtliche Anzahl an Speicherzellen aufaddiert, die für die lokalen Variablen der Funktion belegt werden müssen.
GoTo(name)	Container für ein Goto zu einem anderen Block , wobei das Attribut name der Bezeichner des Labels des Blocks ist zu dem Gesprungen werden soll.
SingleLineComment(prefix, content)	Container für einen Kommentar , den der Compiler selber während des Kompilierungsvorgangs erstellt, der im RETI-Interpreter selbst später nicht sichtbar sein wird, aber in den Immediate-Dateien , welche die Abstrakter Syntaxbaums nach den verschiedenen Passes enthalten.
RETIComment(value)	Container für einen Kommentar im Code der Form: // # comment, der im RETI-Interpreter später sichtbar sein wird und zur Orientierung genutzt werden kann, allerdings in einer tatsächlichen Implementierung einer RETI-CPU nicht umsetzbar ist und auch nicht sinnvoll wäre umzusetzen. Der Kommentar ist im Attribut value, welches jeder Knoten besitzt gespeichert.

Definition 1.8: Label

Durch einen *Bezeichner eindeutig* zuordenbares *Sprungziel* im Programmcode.^a

^aThiemann, „Compilerbau“.

Die ausgegrauten Attribute der PicoC-Nodes sind versteckte Attribute, die **nicht** direkt bei der Erstellung der **PicoC-Nodes** mit einem Wert **initialisiert** werden, sondern im **Verlauf der Kompilierung** beim Durchlaufen der verschiedenen Passes etwas zugewiesen bekommen, dass im weiteren Kompilierungsvorgang **Informationen** transportiert, die später im Kompilierungsvorgang nicht mehr so leicht zugänglich wären.

Jeder **Knoten** hat darüberhinaus auch noch 2 **Attribute** **value** und **position**, wobei **value** bei einem **Token-Knoten** (Definition 1.9) dem **Tokenwert** des Tokens, welches es ersetzt entspricht und bei **Container-Knoten** (Definition 1.10) unbesetzt ist. Das **Attribut position** wird später für Fehlermeldungen gebraucht.

Definition 1.9: Token-Knoten

Ersetzt ein *Token* bei der Generierung des *Abstrakter Syntaxbaum*, damit der Zugriff auf Knoten des Abstrakter Syntaxbaum möglichst *simpel* ist und keine vermeidbaren Fallunterscheidungen gemacht werden müssen.

Token-Knoten entsprechen im Abstrakter Syntaxbaum *Blättern*.^a

^aThiemann, „Compilerbau“.

Definition 1.10: Container-Knoten

Dient als *Container* für andere *Container-Knoten* und *Token-Knoten*. Die *Container-Knoten* werden optimalerweise immer so gewählt, dass sie *mehrere Produktionen der Konkretten Syntax* abdecken, die einen *gleichen Aufbau* haben und sich auch unter einem *Überbegriff* zusammenfassen lassen.^a

Container-Knoten entsprechen im Abstrakter Syntaxbaum *Inneren Knoten*.^b

^aWie z.B. die verschiedenen **Arithmetischen Ausdrücke**, wie z.B. **1 % 3** und **Logischen Ausdrücke**, wie z.B. **1 && 2 < 3**, die einen gleichen Aufbau haben mit immer einer **Operation** in der Mitte haben und 2 **Operanden** auf beiden Seiten und sich unter dem Überbegriff **Binäre Operationen** zusammenfassen lassen.

^bThiemann, „Compilerbau“.

1.2.5.2 RETI-Knoten

Bei den **RETI-Knoten** handelt es sich um Knoten, die irgendeinen **Ausdruck** aus der Sprache L_{RETI} darstellen. Für die **RETI-Knoten** wurden aus bereits in Unterkapitel 1.2.5.1 erläuterten Grund, genauso wie für die **RETI-Knoten** möglichst **kurze** und **leicht** verständliche Bezeichner gewählt. Alle **RETI-Knoten**, die in den von den verschiedenen Passes generierten **Abstrakter Syntaxbaums** vorkommen sind in Tabelle 1.2.5.1 mit einem **Beschreibungstext** dokumentiert.

RETI-Knoten	Beschreibung
Program(name, instrs)	Container für alle Instructions : <name> <instrs>, wobei name der Dateiname der Datei ist, die erstellt wird und instrs eine Liste von Instructions ist.
Instr(op, args)	Container für eine Instruction : <op> <args>, wobei op eine Operation ist und args eine Liste von Argumenten für dieser Operation.
Jump(rel, im_goto)	Container für eine Jump-Instruction : JUMP<rel> <im>, wobei rel eine Relation ist und im_goto ein Immediate Value Im(val) für die Anzahl an Speicherzellen , um die relativ zur Jump-Instruction gesprungen werden soll oder ein GoTo(Name('block.xyz')) , das später im RETI-Patch Pass durch einen passenden Immediate Value ersetzt wird.
Int(num)	Container für einen Interruptaufruf : INT <im>, wobei num die Interruptvektornummer (IVN) für die passende Speicherzelle in der Interruptvektortabelle ist, in der die Adresse der Interrupt-Service-Routine (ISR) steht.
Call(name, reg)	Container für einen Prozeduraufruf : CALL <name> <reg>, wobei name der Bezeichner der Prozedur, die aufgerufen werden soll ist und reg ein Register ist, das als Argument an die Prozedur dient. Diese Operation ist in der Betriebssysteme Vorlesung ^a nicht deklariert, sondern wurde dazuerfunden, um unkompliziert ein CALL PRINT ACC oder CALL INPUT ACC im RETI-Interpreter simulieren zu können.
Name(val)	Bezeichner für eine Prozedur , z.B. PRINT oder INPUT oder den Programmnamen , z.B. PROGRAMNAME. Dieses Argument ist in der Betriebssysteme Vorlesung ^a nicht deklariert, sondern wurde dazuerfunden, um Bezeichner, wie PRINT, INPUT oder PROGRAMNAME schreiben zu können.
Reg(reg)	Container für ein Register .
Im(val)	Ein Immediate Value , z.B. 42, -3 usw.
Add(), Sub(), Mult(), Div(), Mod(), Oplus(), Or(), And()	Compute-Memory oder Compute-Register Operationen: ADD, SUB, MULT, DIV, OPLUS, OR, AND.
Addi(), Subi(), Multi(), Divi(), Modi(), Oplusi(), Ori(), Andi()	Compute-Immediate Operationen: ADDI, SUBI, MULTI, DIVI, MODI, OPLUSI, ORI, ANDI.
Load(), Loadin(), Loadi()	Load Operationen: LOAD, LOADIN, LOADI.
Store(), Storein(), Move()	Store Operationen: STORE, STOREIN, MOVE.
Lt(), LtE(), Gt(), GtE(), Eq(), NEq(), Always(), NOP()	Relationen : <, <=, >, >=, ==, !=, _NOP.
Rti()	Return-From-Interrupt Operation: RTI.
Pc(), In1(), In2(), Acc(), Sp(), Baf(), Cs(), Ds()	Register : PC, IN1, IN2, ACC, SP, BAF, CS, DS.

^a Scholl, „Betriebssysteme“

Tabelle 1.7: RETI-Knoten

1.2.5.3 Kompositionen von PicoC-Knoten und RETI-Knoten mit besonderer Bedeutung

In Tabelle 1.8 sind jegliche **Kompositionen** von **PicoC-Knoten** und **RETI-Knoten** aufgelistet, die eine **besondere Bedeutung** haben und nicht bereits in der **Abstrakten Syntax 1.2.8** enthalten sind.

Komposition	Beschreibung
<code>Ref(Global(Num('addr')))</code>	Speichert Adresse der Speicherzelle, die <code>Num('addr')</code> Speicherzellen relativ zum Datensegment Register DS steht auf den Stack .
<code>Ref(Stackframe(Num('addr')))</code>	Speichert Adresse der Speicherzelle, die <code>Num('addr')</code> Speicherzellen relativ zum Begin-Aktive-Funktion Register BAF steht auf den Stack .
<code>Ref(Subscr(Stack(Num('addr1')), Stack(Num('addr2'))))</code>	Berechnet die nächste Adresse aus der Adresse , die an Speicherzelle <code>Stack(Num('addr1'))</code> steht und dem Subscript Index , der an Speicherzelle <code>Stack(Num('addr2'))</code> steht und speichert diese auf den Stack . Die Berechnung ist abhängig davon ob der Datentyp <code>ArrayDecl(datatype)</code> oder <code>PntrDecl(datatype)</code> ist. Der Datentyp ist ein verstecktes Attribut von <code>Ref(exp)</code> .
<code>Ref(Attr(Stack(Num('addr1')), Name('attr')))</code>	Berechnet die nächste Adresse aus der Adresse , die an Speicherzelle <code>Stack(Num('addr1'))</code> steht und dem Attributnamen <code>Name('attr')</code> und speichert diese auf den Stack . Zur Berechnung ist der Name des Struct in <code>StructSpec(Name('st'))</code> notwendig, dessen Attribut <code>Name('attr')</code> ist. <code>StructSpec(Name('st'))</code> ist ein verstecktes Attribut von <code>Ref(exp)</code> .
<code>Assign(Stack(Num('size')), Global(Num('addr')))</code>	Schreibt <code>Num('size')</code> viele Speicherzellen, die ab <code>Global(Num('addr'))</code> relativ zum Datensegment Register DS stehen, versetzt genauso auf den Stack .
<code>Assign(Stack(Num('size')), Stackframe(Num('addr')))</code>	Schreibt <code>Num('size')</code> viele Speicherzellen, die ab <code>Stackframe(Num('addr'))</code> relativ zum Begin-Aktive-Funktion Register BAF stehen, versetzt genauso auf den Stack .
<code>Exp(Global(Num('addr')))</code>	Speichert Inhalt der Speicherzelle, die <code>Num('addr')</code> Speicherzellen relativ zum Datensegment Register DS steht auf den Stack .
<code>Exp(Stackframe(Num('addr')))</code>	Speichert Inhalt der Speicherzelle, die <code>Num('addr')</code> Speicherzellen relativ zum Begin-Aktive-Funktion Register BAF steht auf den Stack .
<code>Exp(Stack(Num('addr')))</code>	Speichert Inhalt der Speicherzelle, die <code>Num('addr')</code> Speicherzellen relativ zum Stackpointer Register SP steht auf den Stack .
<code>Assign(Stack(Num('addr1')), Stack(Num('addr2')))</code>	Speichert Inhalt der Speicherzelle <code>Stack(Num('addr2'))</code> , die <code>Num('addr2')</code> Speicherzellen relativ zum Stackpointer Register SP steht an der Adresse in der Speicherzelle, die <code>Num('addr1')</code> Speicherzellen relativ zum Stackpointer Register SP steht.
<code>Assign(Global(Num('addr')), Stack(Num('size')))</code>	Schreibt <code>Num('size')</code> viele Speicherzellen, die auf dem Stack stehen, versetzt genauso auf die Speicherzellen ab <code>Num('addr')</code> relativ zum Datensegment Register DS .
<code>Assign(Stackframe(Num('addr')), Stack(Num('size')))</code>	Schreibt <code>Num('size')</code> viele Speicherzellen, die auf dem Stack stehen, versetzt genauso auf die Speicherzellen ab <code>Num('addr')</code> relativ zum Begin-Aktive-Funktion Register BAF .
<code>Exp(Reg(reg))</code>	Schreibt den aktuellen Wert des Registers <code>reg</code> auf den Stack .
<code>Instr(Loadi(), [Reg(Acc()), GoTo(Name('addr@next_instr'))])</code>	Lädt in das Register ACC die Adresse der Instruction, die in diesem Kontext direkt nach dem Sprung zum Block einer anderen Funktion steht.

Tabelle 1.8: Kompositionen von PicoC-Knoten und RETI-Knoten mit besonderer Bedeutung

Um die obige Tabelle 1.8 nicht mit unnötig viel repetitiven Inhalt zu füllen, wurden die zahlreichen Kompositionen **ausgelassen**, bei denen einfach nur **exp** durch **Stack(Num('x'))**, $x \in \mathbb{N}$ ersetzt wurde.

Zudem sind auch jegliche Kombinationen ausgelassen, bei denen einfach nur eine **Expression** an ein **Exp(exp)** bzw. **Ref(exp)** drangehängt wurde.

1.2.5.4 Abstrakte Syntax

Die **Abstrakte Syntax** der Sprache L_{PicoC} wird durch die Grammatik 1.2.10 beschrieben.

<i>stmt</i>	::=	<i>SingleLineComment(str, str)</i> <i>RETIComment()</i>	<i>L_Comment</i>
<i>un_op</i>	::=	<i>Minus()</i> <i>Not()</i>	<i>L_Arith</i>
<i>bin_op</i>	::=	<i>Add()</i> <i>Sub()</i> <i>Mul()</i> <i>Div()</i> <i>Mod()</i> <i>Oplus()</i> <i>And()</i> <i>Or()</i>	
<i>exp</i>	::=	<i>Name(str)</i> <i>Num(str)</i> <i>Char(str)</i> <i>BinOp(<exp>, <bin_op>, <exp>)</i> <i>UnOp(<un_op>, <exp>)</i> <i>Call(Name('input'), Empty())</i> <i>Call(Name('print'), <exp>)</i>	
<i>stmt</i>	::=	<i>Exp(<exp>)</i>	
<i>un_op</i>	::=	<i>LogicNot()</i>	<i>L_Logic</i>
<i>rel</i>	::=	<i>Eq()</i> <i>NEq()</i> <i>Lt()</i> <i>LtE()</i> <i>Gt()</i> <i>GtE()</i>	
<i>bin_op</i>	::=	<i>LogicAnd()</i> <i>LogicOr()</i>	
<i>exp</i>	::=	<i>Atom(<exp>, <rel>, <exp>)</i> <i>ToBool(<exp>)</i>	
<i>type_qual</i>	::=	<i>Const()</i> <i>Writeable()</i>	<i>L_Assign_Alloc</i>
<i>datatype</i>	::=	<i>IntType()</i> <i>CharType()</i> <i>VoidType()</i>	
<i>exp</i>	::=	<i>Alloc(<type_qual>, <datatype>, Name(str))</i>	
<i>stmt</i>	::=	<i>Assign(<exp>, <exp>)</i>	
<i>datatype</i>	::=	<i>PntrDecl(Num(str), <datatype>)</i>	<i>L_Pntr</i>
<i>exp</i>	::=	<i>Deref(<exp>, <exp>)</i> <i>Ref(<exp>)</i>	
<i>datatype</i>	::=	<i>ArrayDecl(Num(str)+, <datatype>)</i>	<i>L_Array</i>
<i>exp</i>	::=	<i>Subscr(<exp>, <exp>)</i> <i>Array(<exp>+)</i>	
<i>datatype</i>	::=	<i>StructSpec(Name(str))</i>	<i>L_Struct</i>
<i>exp</i>	::=	<i>Attr(<exp>, Name(str))</i> <i>StructAssign(Name(str), <exp>+)</i>	
<i>decl_def</i>	::=	<i>StructDecl(Name(str),</i> <i>Alloc(Writeable(), <datatype>, Name(str))+)</i>	
<i>stmt</i>	::=	<i>If(<exp>, <stmt>*)</i> <i>IfElse(<exp>, <stmt>*, <stmt>*)</i>	<i>L_If_Else</i>
<i>stmt</i>	::=	<i>While(<exp>, <stmt>*)</i> <i>DoWhile(<exp>, <stmt>*)</i>	<i>L_Loop</i>
<i>exp</i>	::=	<i>Call(Name(str), <exp>*)</i>	<i>L_Fun</i>
<i>stmt</i>	::=	<i>Return(<exp>)</i>	
<i>decl_def</i>	::=	<i>FunDecl(<datatype>, Name(str),</i> <i>Alloc(Writeable(), <datatype>, Name(str))*)</i> <i>FunDef(<datatype>, Name(str),</i> <i>Alloc(Writeable(), <datatype>, Name(str))* <stmt>*)</i>	
<i>file</i>	::=	<i>File(Name(str), <decl_def>*)</i>	<i>L_File</i>

Grammatik 1.2.10: Abstrakte Syntax der Sprache L_{PicoC}

Man spricht hier von der „**Abstrakten Syntax der Sprache L_{PicoC}** “ und meint hier mit der Sprache L_{PicoC} **nicht** die Sprache, welche durch die **Abstrakte Syntax** beschrieben wird. Es ist damit **immer** die Sprache gemeint, die **kompiliert** werden soll und zu deren Zweck die **Abstrakt Syntax** überhaupt definiert wird. Für die tatsächliche Sprache, die durch die **Abstrakt Syntax** beschrieben wird, interessiert man sich nie wirklich explizit. Diese **Redeart** wurde aus der **Quelle** G. Siek, *Course Webpage for Compilers (P423, P523, E313, and E513)* übernommen.

1.2.5.5 Codebeispiel

In Code 1.5 ist der **Abstrakter Syntaxbaum** zu sehen, der aus dem **vereinfachten Ableitungsbaum** aus Code 1.4 mithilfe eines **Transformers** generiert wurde.

```

1 File
2   Name './verbose_dt_simple_ast_gen_array_decl_and_alloc.ast',
3   [
4     StructDecl
5       Name 'st',
6       [
7         Alloc
8           Writeable,
9           PtrDecl
10            Num '1',
11            ArrayDecl
12              [
13                Num '4',
14                Num '5'
15              ],
16            PtrDecl
17              Num '1',
18              IntType 'int',
19            Name 'attr'
20          ],
21      FunDef
22        VoidType 'void',
23        Name 'main',
24        [],
25        [
26          Exp
27            Alloc
28              Writeable,
29              ArrayDecl
30                [
31                  Num '3',
32                  Num '2'
33                ],
34              PtrDecl
35                Num '1',
36              PtrDecl
37                Num '1',
38              StructSpec
39                Name 'st',
40              Name 'var'
41            ]
42          ]

```


Code 1.5: *Aus vereinfachtem Ableitungsbaum generierter Abstrakter Syntaxbaum***1.2.5.6 Ausgabe des Abstrakter Syntaxbaum**

Ein **Knoten** eines **Abstrakter Syntaxbaum** kann entweder in der **Konkreter Syntax** der Sprache, für dessen Kompilierung er generiert wurde oder in der **Abstrakter Syntax**, die beschreibt, wie der Abstrakter Syntaxbaum selbst aufgebaut sein darf ausgegeben werden.

Das Ausgeben eines **Abstrakter Syntaxbaums** wird im **PicoC-Compiler** über die **Magische Methode** `__repr__()`¹⁷ der Programmiersprache **Python** umgesetzt. Sobald ein **PicoC-Knoten** oder **RETI-Knoten** ausgegeben werden soll, gibt seine Magische Methode `__repr__()` eine nach der **Abstrakten** oder **Konkreten Syntax** aufgebaute **Textrepräsentation** seiner selbst und all seiner Knoten mit an den richtigen Stellen passend gesetzten **runden öffnenden** (und **schließenden**) **Klammern**, sowie **Kommas** `,`, **Semikolons** `;` usw. zur Darstellung der **Hierarchie** und zur **Abtrennung** zurück. Dabei wird nach dem **Depth-First-Search** Schema der gesamte **Abstract Sybtax Tree** durchlaufen und die Magische `__repr__()`-Methode der verschiedenen Knoten aufgerufen, die immer jeweils die `__repr__()`-Methode ihrer Kinder aufrufen und die zurückgegebene **Textrepräsentation** passend **zusammenfügen** und selbst **zurückgeben**.

Im **PicoC-Compiler** wurden **Abstrakte** und **Konkrete Syntax** miteinander gemischt. Für **PicoC-Knoten** wurde die **Abstrakte Syntax** verwendet, da Passes schließlich auf **Abstrakter Syntaxbaums** operieren. Bei **RETI-Knoten** wurde die **Konkrete Syntax** verwendet, da **Maschinenbefehle** in **Konkreter Syntax** schließlich das **Endprodukt** des Kompiliervorgangs sein sollen. Da die **Abstrakte Syntax** von **RETI-Knoten** so simpel ist, macht es kaum einen Unterschied in der Erkennbarkeit, bis auf fehlende geschiefte Klammern () usw., ob man die **RETI-Knoten** in **Abstrakter** oder **Konkreter Syntax** schreibt. Daher kann man auch einfach gleich die **RETI-Knoten** in **Konkreter Syntax** ausgeben und muss nicht beim letzten **Pass** daran denken, am Ende die **Konkrete**, statt der **Abstrakten Syntax** für die **RETI-Knoten** auszugeben.

1.3 Code Generierung

Nach der Generierung eines **Abstrakter Syntaxbaum** als Ergebnis der **Lexikalischen** und **Syntaktischen Analyse** in Unterkapitel ??, wird in diesem Kapitel mit den verschiedenen **Kompositionen** von **PicoC-Knoten** und **RETI-Knoten** im **Abstrakter Syntaxbaum** als Basis das gewünschte Endprodukt des **PicoC-Compilers**, der **RETI-Code** generiert.

Man steht nun dem Problem gegenüber einen **Abstrakter Syntaxbaum** der Sprache L_{PicoC} , der durch die **Abstrakte Syntax** in Grammatik 1.2.10 spezifiziert ist in einen entsprechenden **Abstrakter Syntaxbaum** der Sprache L_{RETI} umzuformen. Das ganze lässt sich, wie in Unterkapitel ?? bereits beschrieben vereinfachen, indem man dieses Problem in mehrere **Passes** (Definition ??) herunterbricht.

Beim **PicoC-Compiler** handelt es sich um einen **Cross-Compiler** (Definiton ??). Damit **RETI-Code** erzeugt werden kann, der auf der **RETI-Architektur** läuft, muss erst, wie im **T-Diagramm** (siehe Unterkapitel ??) in Abbildung 1.6 zu sehen ist, der **Python-Code** des **PicoC-Compilers** mittels eines Compilers, der z.B. auf einer $X_{86,64}$ -Architektur laufen könnte zu **Bytecode** kompiliert werden. Dieser **Bytecode** wird dann von der **Python-Virtual-Machine** (PVM) interpretiert, welche wiederum auf einer $X_{86,64}$ -Architektur laufen könnte. Und selbst dieses **T-Diagramm** könnte noch ausführlicher ausgedrückt werden, indem nachgeforscht wird, in welcher Sprache eigentlich die **Python-Virtual-Machine** geschrieben war, bevor sie zu $X_{86,64}$ kompiliert wurde usw.

¹⁷Spezielle Methode, die immer aufgerufen wird, wenn das **Object**, dass in Besitz dieser Methode ist als **String** mittels `print()` oder zur **Repräsentation** ausgegeben werden soll.

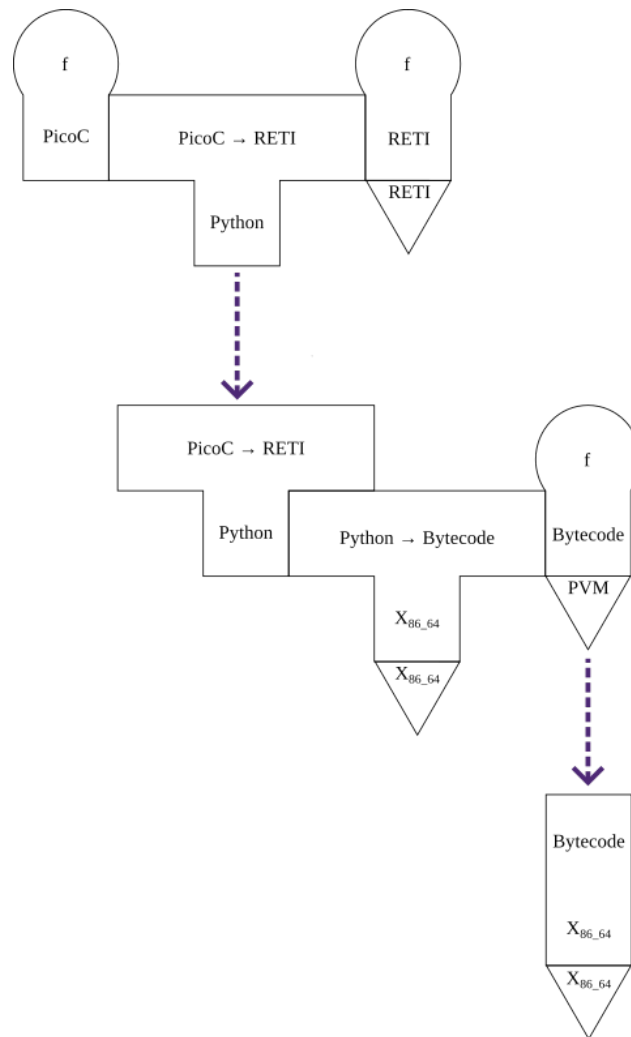


Abbildung 1.6: *Cross-Compiler Kompiliervorgang ausgeschrieben*

Dieses längliche **T-Diagramm** in Abbildung 1.6 lässt sich zusammenfassen, sodass man das **T-Diagramm** in Abbildung 1.7 erhält, in welcher direkt angegeben ist, dass der **PicoC-Compiler** in X_{86_64} -Maschinensprache geschrieben ist.

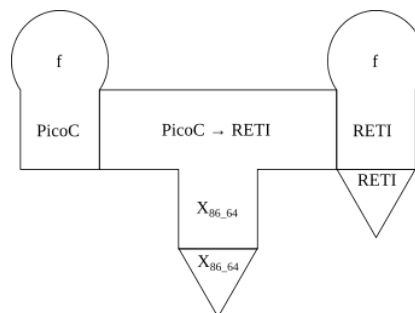


Abbildung 1.7: *Cross-Compiler Kompiliervorgang Kurzform*

Nachdem der Kompilierprozess des **PicoC-Compiler** im **vertikalen** nun genauer angesehen wurde, wird der

Kompilierprozess im Folgenden im **horizontalen**, auf der Ebene der verschiedenen **Passes** genauer betrachtet. Die Abbildung 1.8 gibt einen guten Überblick über alle **Passes** und wie diese in der **Pipe-Architektur** (Definition ??) des **PicoC-Compilers** aufeinanderfolgen. In der **Pipe-Architektur** nutzt der jeweils nächste **Pass** den generierten **Abstrakter Syntaxbaum** des vorherigen Passes oder der Syntaktischen Analyse, um einen eigenen **Abstrakter Syntaxbaum** in seiner eigenen **Sprache** zu generieren.

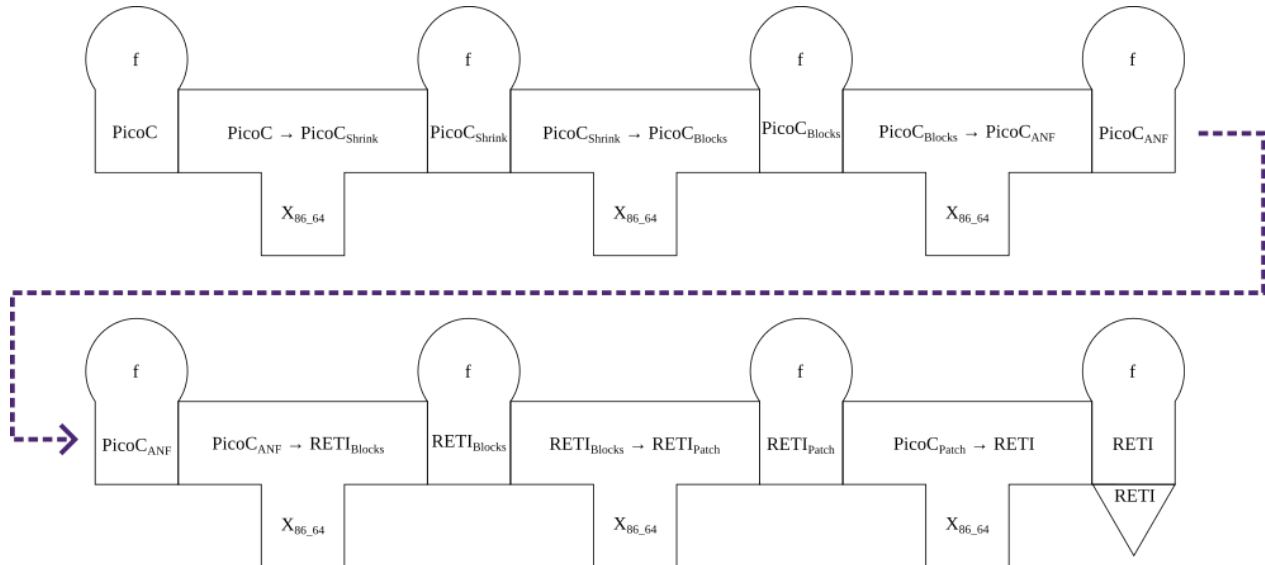


Abbildung 1.8: Architektur mit allen Passes ausgeschrieben

Im Unterkapitel 1.3.1 werden die unterschiedlichen **Passes** des PicoC-Compilers erklärt. In den darauffolgenden Unterkapiteln ??, ??, ?? und ?? zu **Pointern**, **Arrays**, **Structs** und **Funktionen** werden einzelne **Aspekte**, die Thema dieser **Bachelorarbeit** sind **genauer betrachtet** und erklärt, die im Unterkapitel 1.3.1 nicht ausreichend vertieft wurden. Viele der verwendeten **Ansätze** zur Lösung dieser Probleme basieren auf der Vorlesung Scholl, „Betriebssysteme“ und wurden in dieser Bachelorarbeit weiter ausgearbeitet, wo es nötig war, sodass diese mit dem **PicoC-Compiler** auch in der **Praxis** implementiert werden konnten.

Um die verschiedenen Aspekte besser erklären zu können, werden **Codebeispiele** verwendet, in welchen ein kleines repräsentatives **PicoC-Programm** für einen spezifischen Aspekt in wichtigen **Zwischenstadien der Kompilierung** gezeigt wird¹⁸. Die **Codebeispiele** wurden alle mit dem **PicoC-Compiler** kompiliert und danach **nicht** mehr **verändert**, also genauso, wie der **PicoC-Compiler** sie kompiliert aus den Dateien in dieses Dokument eingelesen. Alle hier zur Repräsentation verwendeten **PicoC-Programme** lassen sich unter dem **Link**¹⁹ finden und mithilfe der im Ordner `/code_examples` beiliegenden `Makefile` und dem Befehl `> make compile-all` genauso **kompilieren**, wie sie hier dargestellt sind²⁰.

1.3.1 Passes

Im Folgenden werden die verschiedenen **Passes** des **PicoC-Compilers** für die Generierung von **RETI-Code** besprochen. Viele dieser **Passes** haben **Aufgaben**, die eher unter die Themenbereiche des **Bachelorprojekts** fallen. Allerdings ist das Verständnis der **Passes** auch für das Verständnis der verschiedenen Aspekte²¹ der

¹⁸Also die verschiedenen in den **Passes** generierten **Abstrakter Syntaxbaums**, sofern der **Pass** für den gezeigten Aspekt relevant ist.

¹⁹https://github.com/matthejue/Bachelorarbeit/tree/master/code_examples.

²⁰Es wurden zu diesem Zweck spezielle neue **Command-line Optionen** erstellt, die bestimmte Kommentare **herausfiltern** und manche Container-Knoten **einzeilig** machen, damit die generierten **Abstrakter Syntaxbaums** in den verschiedenen Zwischenstufen der Kompilierung **nicht** zu langgestreckt und **überfüllt** mit Kommentaren sind.

²¹In kurz: **Pointer**, **Arrays**, **Structs** und **Funktionen**.

Bachelorarbeit wichtig.

Auf jedes Detail der einzelnen **Passes** wird in diesem Unterkapitel allerdings nicht eingegangen, da diese einerseits in den Unterkapiteln ??, ??, ?? und ?? zu **Pointern**, **Arrays**, **Structs** und **Funktionen** im Detail erklärt sind und andererseits viele Aufgaben dieser **Passes** eher dem **Bachelorprojekt** zuzurechnen sind.

1.3.1.1 PicoC-Shrink Pass

1.3.1.1.1 Aufgabe

Der Aufgabe des **PicoC-Shrink Pass** ist in Unterkapitel ?? ausführlich an einem Beispiel erklärt. Kurzgefasst hat der **PicoC-Shrink Pass** die Aufgabe, die Eigenheit auszunutzen, dass der **Dereferenzierungoperator** ***pntr** und die damit einhergehende **Pointer Arithmetik** ***(pntr + i)** sich in der Untermenge der Sprache L_C , welche die Sprache L_{PicoC} darstellt genau gleich verhält, wie der **Operator** für den **Zugriff** auf den **Index** eines **Arrays** **ar[i]**.

Daher wandelt der **PicoC-Shrink Pass** alle Verwendungen des **Knoten** **Deref(exp, i)** im jeweiligen **Abstrakter Syntaxbaum** in **Knoten** **Subscr(exp, i)** um, sodass sich dadurch viele vermeidbare **Fallunterscheidungen** und **doppelter Code** bei der Implementierung vermeiden lassen. Man lässt die **Dereferenzierung** ***(var + i)** einfach von den Routinen für einen **Zugriff auf einen Arrayindex** **var[i]** übernehmen.

1.3.1.1.2 Abstrakte Syntax

Die **Abstrakte Syntax** der Sprache L_{PicoC_Shrink} in Tabelle 1.3.1 ist fast **identisch** mit der **Abstrakten Syntax** der Sprache L_{PicoC} in Tabelle 1.2.10, nach welcher der **erste** Abstrakter Syntaxbaum in der **Syntaktischen Analyse** generiert wurde. Der einzige **Unterschied** liegt darin, dass es den Knoten **Deref(exp, exp)** in Tabelle 1.3.1 **nicht** mehr gibt. Das liegt daran, dass dieser Pass alle **Vorkommnisse** des Knoten **Deref(exp, exp)** durch den Knoten **Subscr(exp, exp)** auswechselt, der ebenfalls bereits in der **Abstrakten Syntax** der Sprache L_{PicoC} definiert ist.

<i>stmt</i>	::=	<i>SingleLineComment</i> (<i>str</i> , <i>str</i>) <i>RETIComment</i> ()	<i>L_Comment</i>
<i>un_op</i>	::=	<i>Minus</i> () <i>Not</i> ()	<i>L_Arith</i>
<i>bin_op</i>	::=	<i>Add</i> () <i>Sub</i> () <i>Mul</i> () <i>Div</i> () <i>Mod</i> () <i>Oplus</i> () <i>And</i> () <i>Or</i> ()	
<i>exp</i>	::=	<i>Name</i> (<i>str</i>) <i>Num</i> (<i>str</i>) <i>Char</i> (<i>str</i>) <i>BinOp</i> (<i><exp></i> , <i><bin_op></i> , <i><exp></i>) <i>UnOp</i> (<i><un_op></i> , <i><exp></i>) <i>Call</i> (<i>Name</i> ('input'), <i>Empty</i> ()) <i>Call</i> (<i>Name</i> ('print'), <i><exp></i>)	
<i>stmt</i>	::=	<i>Exp</i> (<i><exp></i>)	
<i>un_op</i>	::=	<i>LogicNot</i> ()	<i>L_Logic</i>
<i>rel</i>	::=	<i>Eq</i> () <i>NEq</i> () <i>Lt</i> () <i>LtE</i> () <i>Gt</i> () <i>GtE</i> ()	
<i>bin_op</i>	::=	<i>LogicAnd</i> () <i>LogicOr</i> ()	
<i>exp</i>	::=	<i>Atom</i> (<i><exp></i> , <i><rel></i> , <i><exp></i>) <i>ToBool</i> (<i><exp></i>)	
<i>type_qual</i>	::=	<i>Const</i> () <i>Writeable</i> ()	<i>L_Assign_Alloc</i>
<i>datatype</i>	::=	<i>IntType</i> () <i>CharType</i> () <i>VoidType</i> ()	
<i>exp</i>	::=	<i>Alloc</i> (<i><type_qual></i> , <i><datatype></i> , <i>Name</i> (<i>str</i>))	
<i>stmt</i>	::=	<i>Assign</i> (<i><exp></i> , <i><exp></i>)	
<i>datatype</i>	::=	<i>PntrDecl</i> (<i>Num</i> (<i>str</i>), <i><datatype></i>)	<i>L_Pntr</i>
<i>exp</i>	::=	<i>Deref</i> (<i><exp></i> , <i><exp></i>) <i>Ref</i> (<i><exp></i>)	
<i>datatype</i>	::=	<i>ArrayDecl</i> (<i>Num</i> (<i>str</i>) +, <i><datatype></i>)	<i>L_Array</i>
<i>exp</i>	::=	<i>Subscr</i> (<i><exp></i> , <i><exp></i>) <i>Array</i> (<i><exp></i> +)	
<i>datatype</i>	::=	<i>StructSpec</i> (<i>Name</i> (<i>str</i>))	<i>L_Struct</i>
<i>exp</i>	::=	<i>Attr</i> (<i><exp></i> , <i>Name</i> (<i>str</i>)) <i>StructAssign</i> (<i>Name</i> (<i>str</i>), <i><exp></i>) +)	
<i>decl_def</i>	::=	<i>StructDecl</i> (<i>Name</i> (<i>str</i>), <i>Alloc</i> (<i>Writeable</i> (), <i><datatype></i> , <i>Name</i> (<i>str</i>)) +)	
<i>stmt</i>	::=	<i>If</i> (<i><exp></i> , <i><stmt></i> *) <i>IfElse</i> (<i><exp></i> , <i><stmt></i> *, <i><stmt></i> *)	<i>L_If_Else</i>
<i>stmt</i>	::=	<i>While</i> (<i><exp></i> , <i><stmt></i> *) <i>DoWhile</i> (<i><exp></i> , <i><stmt></i> *)	<i>L_Loop</i>
<i>exp</i>	::=	<i>Call</i> (<i>Name</i> (<i>str</i>), <i><exp></i> *)	<i>L_Fun</i>
<i>stmt</i>	::=	<i>Return</i> (<i><exp></i>)	
<i>decl_def</i>	::=	<i>FunDecl</i> (<i><datatype></i> , <i>Name</i> (<i>str</i>), <i>Alloc</i> (<i>Writeable</i> (), <i><datatype></i> , <i>Name</i> (<i>str</i>))*) <i>FunDef</i> (<i><datatype></i> , <i>Name</i> (<i>str</i>), <i>Alloc</i> (<i>Writeable</i> (), <i><datatype></i> , <i>Name</i> (<i>str</i>))*, <i><stmt></i> *)	
<i>file</i>	::=	<i>File</i> (<i>Name</i> (<i>str</i>), <i><decl_def></i> *)	<i>L_File</i>

Grammatik 1.3.1: Abstrakte Syntax der Sprache *L_{PiocC-Shrink}*

Der **rot** markierte Knoten bedeutet, dass dieser im Vergleich zur vorherigen **Abstrakten Syntax** nicht mehr da ist.

1.3.1.1.3 Codebeispiel

In den nächsten Unterkapiteln wird das Beispiel in Code 1.6 zur **Anschauung** der verschiedenen **Passes** verwendet. Im Code 1.6 ist in der Funktion **faculty** ein **iterativer** Algorithmus implementiert, der die **Fakultät** eines übergebenen **Arguments** berechnet. Der Algorithmus basiert auf einem **Beispielprogramm**

aus der Vorlesung Scholl, „Betriebssysteme“, welcher in der Vorlesung allerdings **rekursiv** implementiert ist.

Dieser **rekursive** Algorithmus ist allerdings **kein** gutes **Anschauungsbeispiel**, dass viele der Aufgaben der verschiedenen **Passes** bei der Kompilierung veranschaulicht hätte. Viele Aufgaben der **Passes**, wie z.B. bei der Kompilierung von **if**-, **if-else**-, **while**- und **do-while**-Statements wären im Beispiel aus der Vorlesung **nicht** enthalten gewesen. Daher wurde das Beispiel aus der Vorlesung zu einem **iterativen** Algorithmus 1.6 umgeschrieben, um **if**- und **while**-Statemens zu enthalten.

Beide Varianten des **Algorithmus** wurden zum **Testen** des PicoC-Compilers verwendet und sind als Tests im Ordner `/tests` unter [Link²²](#), unter den Testbezeichnungen `example_faculty_rec.picoc` und `example_faculty_it.picoc` zu finden.

Die Codebeispiele in diesem und den folgenden Unterkapiteln dienen allerdings nur als **Anschauung** des jeweiligen **Passes**, der in diesem Unterkapitel beschrieben wird und werden nicht im Detail erläutert, da viele Details der Passes später in den Unterkapiteln ??, ??, ?? und ?? zu **Pointern**, **Arrays**, **Structs** und **Funktionen** mit eigenen **Codebeispielen** erklärt werden und alle sonstigen Details dem **Bachelorprojekt** zuzurechnen sind.

```

1 // based on a example program from Christoph Scholl's Operating Systems lecture
2
3 int faculty(int n){
4     int res = 1;
5     while (1) {
6         if (n == 1) {
7             return res;
8         }
9         res = n * res;
10        n = n - 1;
11    }
12 }
13
14 void main() {
15     print(faculty(4));
16 }

```

Code 1.6: PicoC Code für Codebespiel

In Code 1.7 sieht man den **Abstrakter Syntaxbaum**, der in der **Syntaktischen Analyse** generiert wurde.

```

1 File
2   Name './example_faculty_it.ast',
3   [
4     FunDef
5       IntType 'int',
6       Name 'faculty',
7       [
8         Alloc(Writable(), IntType('int'), Name('n'))
9       ],

```

²²https://github.com/matthejue/PicoC-Compiler/tree/new_architecture/tests.

```

10  [
11    Assign(Alloc(Writable(), IntType('int'), Name('res')), Num('1')),
12    While
13      Num '1',
14      [
15        If
16          Atom(Name('n'), Eq('=='), Num('1')),
17          [
18            Return(Name('res'))
19          ]
20          Assign(Name('res'), BinOp(Name('n'), Mul('*'), Name('res')))
21          Assign(Name('n'), BinOp(Name('n'), Sub('-'), Num('1')))
22      ]
23  ],
24  FunDef
25    VoidType 'void',
26    Name 'main',
27    [],
28    [
29      Exp(Call(Name('print'), [Call(Name('faculty'), [Num('4')])]))
30    ]
31 ]

```

Code 1.7: Abstrakter Syntaxbaum für Codebeispiel

Im **PicoC-Shrink-Pass** ändert sich nichts im Vergleich zum **Abstrakter Syntaxbaum** in Code 1.7, da das Codebeispiel keine **Dereferenzierung** enthält.

1.3.1.2 PicoC-Blocks Pass

1.3.1.2.1 Aufgabe

Die Aufgabe des **PicoC-Blocks Passes** ist es die Knoten `If(exp, stmts)`, `IfElse(exp, stmts1, stmts2)`, `While(exp, stmts)` und `DoWhile(exp, stmts)` mithilfe von `Block(name, stmts_instrs-)`, `GoTo(label)-` und `IfElse(exp, stmts1, stmts2)-Knoten` umzusetzen. Der `IfElse(exp, stmts1, stmts2)-Knoten` wird zur Umsetzung der **Bedingung** verwendet und es wird, je nachdem, ob die Bedingung **wahr** oder **falsch** ist mithilfe der `GoTo(label)-Knoten` in einen von zwei **alternativen Branches** gesprungen oder ein **Branch** erneut aufgerufen usw.

1.3.1.2.2 Abstrakte Syntax

Zur Umsetzung dieses Passes ist es notwendig die **Abstrakte Syntax** der Sprache L_{PicoC_Shrink} in Tabelle 1.3.1 um die Knoten zu erweitern, die im Unterkapitel 1.3.1.2.1 erwähnt wurden. Die Knoten `If(exp, stmts)`, `While(exp, stmts)` und `DoWhile(exp, stmts)` gibt es nicht mehr, da sie durch `Block(name, stmts_instrs-)`, `GoTo(label)-` und `IfElse(exp, stmts1, stmts2)-Knoten` ersetzt wurden. Die **Funktionsdefinition** `FunDef(<datatype>, Name(str), Alloc(Writable(), <datatype>, Name(str))* , <block>*)` ist nun ein Container für **Blöcke** `Block(Name(str), <stmt>*)` und keine Statements `stmt` mehr. Das resultiert in der **Abstrakten Syntax** der Sprache L_{PicoC_Blocks} in Tabelle 1.3.2.

<i>stmt</i>	::=	<i>SingleLineComment(str, str)</i> <i>RETIComment()</i>	<i>L_Comment</i>
<i>un_op</i>	::=	<i>Minus()</i> <i>Not()</i>	<i>L_Arith</i>
<i>bin_op</i>	::=	<i>Add()</i> <i>Sub()</i> <i>Mul()</i> <i>Div()</i> <i>Mod()</i> <i>Oplus()</i> <i>And()</i> <i>Or()</i>	
<i>exp</i>	::=	<i>Name(str)</i> <i>Num(str)</i> <i>Char(str)</i> <i>BinOp(<exp>, <bin_op>, <exp>)</i> <i>UnOp(<un_op>, <exp>)</i> <i>Call(Name('input'), Empty())</i> <i>Call(Name('print'), <exp>)</i>	
<i>stmt</i>	::=	<i>Exp(<exp>)</i>	
<i>un_op</i>	::=	<i>LogicNot()</i>	<i>L_Logic</i>
<i>rel</i>	::=	<i>Eq()</i> <i>NEq()</i> <i>Lt()</i> <i>LtE()</i> <i>Gt()</i> <i>GtE()</i>	
<i>bin_op</i>	::=	<i>LogicAnd()</i> <i>LogicOr()</i>	
<i>exp</i>	::=	<i>Atom(<exp>, <rel>, <exp>)</i> <i>ToBool(<exp>)</i>	
<i>type_qual</i>	::=	<i>Const()</i> <i>Writeable()</i>	<i>L_Assign_Alloc</i>
<i>datatype</i>	::=	<i>IntType()</i> <i>CharType()</i> <i>VoidType()</i>	
<i>exp</i>	::=	<i>Alloc(<type_qual>, <datatype>, Name(str))</i>	
<i>stmt</i>	::=	<i>Assign(<exp>, <exp>)</i>	
<i>datatype</i>	::=	<i>PntrDecl(Num(str), <datatype>)</i>	<i>L_Pntr</i>
<i>exp</i>	::=	<i>Ref(<exp>)</i>	
<i>datatype</i>	::=	<i>ArrayDecl(Num(str)+, <datatype>)</i>	<i>L_Array</i>
<i>exp</i>	::=	<i>Subscr(<exp>, <exp>)</i> <i>Array(<exp>+)</i>	
<i>datatype</i>	::=	<i>StructSpec(Name(str))</i>	<i>L_Struct</i>
<i>exp</i>	::=	<i>Attr(<exp>, Name(str))</i> <i>StructAssign(Name(str), <exp>)+)</i>	
<i>decl_def</i>	::=	<i>StructDecl(Name(str),</i> <i>Alloc(Writeable(), <datatype>, Name(str))+)</i>	
<i>stmt</i>	::=	<i>If(<exp>, <stmt>*)</i> <i>IfElse(<exp>, <stmt>*, <stmt>*)</i>	<i>L_If_Else</i>
<i>stmt</i>	::=	<i>While(<exp>, <stmt>*)</i> <i>DoWhile(<exp>, <stmt>*)</i>	<i>L_Loop</i>
<i>exp</i>	::=	<i>Call(Name(str), <exp>*)</i>	<i>L_Fun</i>
<i>stmt</i>	::=	<i>Return(<exp>)</i>	
<i>decl_def</i>	::=	<i>FunDecl(<datatype>, Name(str),</i> <i>Alloc(Writeable(), <datatype>, Name(str))*)</i> <i>FunDef(<datatype>, Name(str),</i> <i>Alloc(Writeable(), <datatype>, Name(str))*</i> , <i><block>*)</i>	
<i>block</i>	::=	<i>Block(Name(str), <stmt>*)</i>	<i>L_Blocks</i>
<i>stmt</i>	::=	<i>GoTo(Name(str))</i>	
<i>file</i>	::=	<i>File(Name(str), <decl_def>*)</i>	<i>L_File</i>

Grammatik 1.3.2: Abstrakte Syntax der Sprache L_{Pioc_Blocks}

Alles **rot** markierte bedeutet, es wurde **entfernt** oder **abgeändert**. Alles **ausgegraute** bedeutet, es hat sich im Vergleich zur letzten Abstrakten Syntax **nichts** geändert. Alle normal in **schwarz** geschriebenen Knoten wurden **neu** hinzugefügt.

Die **Abstrakte Syntax** soll im Gegensatz zur **Konkreten Syntax** meist nur vom **Programmierer**

verstanden werden, der den Compiler implementiert und sollte daher vor allem **einfach verständlich** sein und stellt daher eine **Obermenge** aller tatsächlich möglichen **Kompositionen** von **Knoten** dar^a.

Man bezeichnet hier die **Abstrakte Syntax** als „**Abstrakte Syntax der Sprache** L_{Picoc_Blocks} “. Diese Sprache L_{Picoc_Blocks} wird durch eine **Konkrete Syntax** beschrieben, die allerdings nicht weiter relevant ist, da in den **Passes** nur **Abstrakter Syntaxbaums** umgeformt werden. Es ist hierbei nur wichtig zu wissen, dass die **Abstrakte Syntax** theoretisch zur Kompilierung der Sprache L_{Picoc_Blocks} definiert ist, also die Sprache L_{PicoC_Blocks} nicht die Sprache ist, die von der **Abstrakten Syntax** beschrieben ist.

^aD.h. auch wenn dort **exp** als **Attribut** steht, kann dort **nicht** jeder Knoten, der sich aus der **Produktion** **exp** ergibt auch wirklich eingesetzt werden.

1.3.1.2.3 Codebeispiel

In Code 1.8 sieht man den **Abstract-Syntax-Tree** des **PiocC-Blocks Passes** für das aus Unterkapitel 1.6 weitergeführte Beispiel, indem nun eigene **Blöcke** für die Funktion **faculty** und die **main**-Funktion erstellt werden, in denen die **ersten** Statements der jeweiligen Funktionen bis zum **letzten** Statement oder bis zum ersten **Auftauchen** eines **If(exp, stmts)-**, **IfElse(exp, stmts1, stmts2)-**, **While(exp, stmts)-** oder **DoWhile(exp, stmts)-Knoten** stehen. Je nachdem, ob ein **If(exp, stmts)-**, **IfElse(exp, stmts1, stmts2)-**, **While(exp, stmts)-** oder **DoWhile(exp, stmts)-Knoten** auftaucht, werden für die **Bedingung** und mögliche **Branches** eigene **Blöcke** erstellt.

```

1 File
2   Name './example_faculty_it.picoc_blocks',
3   [
4     FunDef
5       IntType 'int',
6       Name 'faculty',
7       [
8         Alloc(Writable(), IntType('int'), Name('n'))
9       ],
10      [
11        Block
12          Name 'faculty.6',
13          [
14            Assign(Alloc(Writable(), IntType('int'), Name('res')), Num('1'))
15            // While(Num('1'), [])
16            GoTo(Name('condition_check.5'))
17          ],
18        Block
19          Name 'condition_check.5',
20          [
21            IfElse
22              Num '1',
23              [
24                GoTo(Name('while_branch.4'))
25              ],
26              [
27                GoTo(Name('while_after.1'))
28              ]
29          ],
30        Block
31          Name 'while_branch.4',

```

```

32     [
33         // If(Atom(Name('n'), Eq('=='), Num('1')), [],),
34         IfElse
35             Atom(Name('n'), Eq('=='), Num('1')),
36             [
37                 GoTo(Name('if.3'))
38             ],
39             [
40                 GoTo(Name('if_else_after.2'))
41             ]
42     ],
43     Block
44         Name 'if.3',
45         [
46             Return(Name('res'))
47         ],
48     Block
49         Name 'if_else_after.2',
50         [
51             Assign(Name('res'), BinOp(Name('n'), Mul('*'), Name('res')))
52             Assign(Name('n'), BinOp(Name('n'), Sub('-'), Num('1')))
53             GoTo(Name('condition_check.5'))
54         ],
55     Block
56         Name 'while_after.1',
57         []
58 ],
59 FunDef
60     VoidType 'void',
61     Name 'main',
62     [],
63     [
64         Block
65             Name 'main.0',
66             [
67                 Exp(Call(Name('print'), [Call(Name('faculty'), [Num('4')])]))
68             ]
69     ]
70 ]

```

Code 1.8: PicoC-Blocks Pass für Codebeispiel

1.3.1.3 PicoC-ANF Pass

1.3.1.3.1 Aufgabe

Die Aufgabe des **PicoC-ANF Passes** ist es den **Abstrakter Syntaxbaum** der Sprache L_{PicoC_Blocks} in die **Abstrakte Syntax** der Sprache L_{PicoC_ANF} umzuformen, welche in **A-Normalform** (Definition ??) und damit auch in **Monadischer Normalform** (Definition ??) ist. Um Wiederholung zu vermeiden wird zur Erklärung der **A-Normalform** auf Unterkapitel ?? verwiesen.

Zudem wird eine **Symboltabelle** (Definition 1.11) eingeführt. In der **Symboltabelle** wird beim Anlegen eines **neuen Eintrags** für eine Variable zunächst eine **Adresse** zugewiesen, die dem Wert einer von zwei **Countern** `rel_global_addr` und `rel_stack_addr` entspricht. Der Counter `rel_global_addr` ist für Variablen in den **Globalen Statischen Daten** und der Counter `rel_stack_addr` ist für Variablen auf dem **Stackframe**. Einer der beiden **Counter** wird entsprechend der **Größe** der angelegten Variable **hochgezählt**.

Kommt im **Programmcode** an einer späteren Stelle diese Variable `Name('symbol')` vor, so wird mit dem **Symbol**²³ als Schlüssel in der **Symbole** nachgeschlagen und anstelle des `Name(str)`-Knotens die in der **Symbole** nachgeschlagene Adresse in einem `Global(Num('addr'))`- bzw. `Stackframe(Num('addr'))`-Knoten eingesetzt eingefügt. Ob der `Global(Num('addr'))`- oder der `Stackframe(Num('addr'))`-Knoten zum Einsatz kommt, entscheidet sich anhand des **Scopes** (z.B. `@scope`), der in der **Symbole** an den **Bezeichner** drangehängt ist (z.B. `identifizier@scope`).²⁴

Definition 1.11: Symbole

*Eine über ein **Assoziatives Feld** umgesetzte **Datenstruktur**, die notwendig ist, um das Konzept einer **Variablen** in einer Sprache umzusetzen. Diese Datenstruktur ordnet jedem **Symbol**^a einer **Variablen**, **Konstanten** oder **Funktion** aus einem **Programm**, Informationen, wie die **Adresse**, die **Position** im Programmcode oder den **Datentyp** zu.*

*Die **Symbole** muss nur während des **Kompilervorgangs** im **Speicher** existieren, da die Einträge in der **Symbole** beeinflussen, was für **Maschinencode** generiert wird und dadurch im **Maschinencode** bereits die richtigen **Adressen** usw. angesprochen werden und es die **Symbole** selbst **nicht** mehr braucht.*

^aIn einer **Symbole** werden **Bezeichner** als **Symbole** bezeichnet.

1.3.1.3.2 Abstrakte Syntax

Zur Umsetzung dieses Passes ist es notwendig die **Abstrakte Syntax** der Sprache L_{PicoC_Blocks} in Tabelle 1.3.2 in die **A-Normalform** zu bringen. Darunter fällt es unter anderem, dafür zu sorgen, dass **Komplexe Knoten**, wie z.B. `BinOp(exp, bin_op, exp)` nur **Atomare Knoten**, wie z.B. `Stack(Num(str))` enthalten können. Des Weiteren werden auch **Funktionen** und **Funktionsaufrufe** aufgelöst, sodass u.a. die **Blöcke** `Block(Name(str), stmt*)` nun direkt im `File(Name(str), block*)`-Knoten liegen usw., was in Unterkapitel ?? genauer erklärt wird. Die **Symbole** ist ebenfalls als **Abstrakter Syntaxbaum** umgesetzt, wofür in der **Abstrakten Syntax** der Sprache L_{PicoC_ANF} in Grammatik 1.3.3 neue Knoten eingeführt werden.

Das ganze resultiert in der **Abstrakten Syntax** der Sprache L_{PicoC_ANF} in Grammatik 1.3.3.

²³Bzw. der **Bezeichner**

²⁴Die Umsetzung von **Scopes** wird in Unterkapitel ?? genauer beschrieben.

<i>stmt</i>	::=	<i>SingleLineComment</i> (<i>str</i> , <i>str</i>) <i>RETIComment</i> ()	<i>L_Comment</i>
<i>un_op</i>	::=	<i>Minus</i> () <i>Not</i> ()	<i>L_Arith</i>
<i>bin_op</i>	::=	<i>Add</i> () <i>Sub</i> () <i>Mul</i> () <i>Div</i> () <i>Mod</i> () <i>Oplus</i> () <i>And</i> () <i>Or</i> ()	
<i>exp</i>	::=	<i>Name</i> (<i>str</i>) <i>Num</i> (<i>str</i>) <i>Char</i> (<i>str</i>) <i>Global</i> (<i>Num</i> (<i>str</i>)) <i>Stackframe</i> (<i>Num</i> (<i>str</i>)) <i>Stack</i> (<i>Num</i> (<i>str</i>)) <i>BinOp</i> (<i>Stack</i> (<i>Num</i> (<i>str</i>)), <i>bin_op</i> , <i>Stack</i> (<i>Num</i> (<i>str</i>))) <i>UnOp</i> (<i>un_op</i> , <i>Stack</i> (<i>Num</i> (<i>str</i>))) <i>Call</i> (<i>Name</i> ('input'), <i>Empty</i> ()) <i>Call</i> (<i>Name</i> ('print'), <i>exp</i>) <i>Exp</i> (<i>exp</i>)	
<i>un_op</i>	::=	<i>LogicNot</i> ()	<i>L_Logic</i>
<i>rel</i>	::=	<i>Eq</i> () <i>NEq</i> () <i>Lt</i> () <i>LtE</i> () <i>Gt</i> () <i>GtE</i> ()	
<i>bin_op</i>	::=	<i>LogicAnd</i> () <i>LogicOr</i> ()	
<i>exp</i>	::=	<i>Atom</i> (<i>Stack</i> (<i>Num</i> (<i>str</i>)), <i>rel</i> , <i>Stack</i> (<i>Num</i> (<i>str</i>))) <i>ToBool</i> (<i>Stack</i> (<i>Num</i> (<i>str</i>)))	
<i>type_qual</i>	::=	<i>Const</i> () <i>Writeable</i> ()	<i>L_Assign_Alloc</i>
<i>datatype</i>	::=	<i>IntType</i> () <i>CharType</i> () <i>VoidType</i> ()	
<i>exp</i>	::=	<i>Alloc</i> (<i>type_qual</i> , <i>datatype</i> , <i>Name</i> (<i>str</i>))	
<i>stmt</i>	::=	<i>Assign</i> (<i>Global</i> (<i>Num</i> (<i>str</i>)), <i>Stack</i> (<i>Num</i> (<i>str</i>))) <i>Assign</i> (<i>Stackframe</i> (<i>Num</i> (<i>str</i>)), <i>Stack</i> (<i>Num</i> (<i>str</i>))) <i>Assign</i> (<i>Stack</i> (<i>Num</i> (<i>str</i>)), <i>Global</i> (<i>Num</i> (<i>str</i>))) <i>Assign</i> (<i>Stack</i> (<i>Num</i> (<i>str</i>)), <i>Stackframe</i> (<i>Num</i> (<i>str</i>)))	
<i>datatype</i>	::=	<i>PntrDecl</i> (<i>Num</i> (<i>str</i>), <i>datatype</i>) <i>Ref</i> (<i>Global</i> (<i>str</i>)) <i>Ref</i> (<i>Stackframe</i> (<i>str</i>)) <i>Ref</i> (<i>Subscr</i> (<i>exp</i> , <i>exp</i>) <i>Ref</i> (<i>Attr</i> (<i>exp</i> , <i>Name</i> (<i>str</i>))))	<i>L_Pntr</i>
<i>datatype</i>	::=	<i>ArrayDecl</i> (<i>Num</i> (<i>str</i>)+, <i>datatype</i>)	<i>L_Array</i>
<i>exp</i>	::=	<i>Subscr</i> (<i>exp</i> , <i>Stack</i> (<i>Num</i> (<i>str</i>))) <i>Array</i> (<i>exp</i>)+)	
<i>datatype</i>	::=	<i>StructSpec</i> (<i>Name</i> (<i>str</i>))	<i>L_Struct</i>
<i>exp</i>	::=	<i>Attr</i> (<i>exp</i> , <i>Name</i> (<i>str</i>)) <i>Struct</i> (<i>Assign</i> (<i>Name</i> (<i>str</i>), <i>exp</i>))+)	
<i>decl_def</i>	::=	<i>StructDecl</i> (<i>Name</i> (<i>str</i>), <i>Alloc</i> (<i>Writeable</i> (), <i>datatype</i> , <i>Name</i> (<i>str</i>))+)	
<i>stmt</i>	::=	<i>IfElse</i> (<i>Stack</i> (<i>Num</i> (<i>str</i>)), <i>stmt</i> *, <i>stmt</i> *)	<i>L_If_Else</i>
<i>exp</i>	::=	<i>Call</i> (<i>Name</i> (<i>str</i>), <i>exp</i> *)	<i>L_Fun</i>
<i>stmt</i>	::=	<i>StackMalloc</i> (<i>Num</i> (<i>str</i>)) <i>NewStackframe</i> (<i>Name</i> (<i>str</i>), <i>GoTo</i> (<i>str</i>)) <i>Exp</i> (<i>GoTo</i> (<i>Name</i> (<i>str</i>))) <i>RemoveStackframe</i> () <i>Return</i> (<i>Empty</i> ()) <i>Return</i> (<i>exp</i>)	
<i>decl_def</i>	::=	<i>FunDecl</i> (<i>datatype</i> , <i>Name</i> (<i>str</i>) <i>Alloc</i> (<i>Writeable</i> (), <i>datatype</i> , <i>Name</i> (<i>str</i>))* <i>FunDef</i> (<i>datatype</i> , <i>Name</i> (<i>str</i>), <i>Alloc</i> (<i>Writeable</i> (), <i>datatype</i> , <i>Name</i> (<i>str</i>))* <i>block</i> *)	
<i>block</i>	::=	<i>Block</i> (<i>Name</i> (<i>str</i>), <i>stmt</i> *)	<i>L_Blocks</i>
<i>stmt</i>	::=	<i>GoTo</i> (<i>Name</i> (<i>str</i>))	
<i>file</i>	::=	<i>File</i> (<i>Name</i> (<i>str</i>), <i>block</i> *)	<i>L_File</i>
<i>symbol_table</i>	::=	<i>SymbolTable</i> (<i>symbol</i> *)	<i>L_Symbol_Table</i>
<i>symbol</i>	::=	<i>Symbol</i> (<i>type_qual</i> , <i>datatype</i> , <i>name</i> , <i>val</i> , <i>pos</i> , <i>size</i>)	
<i>type_qual</i>	::=	<i>Empty</i> ()	
<i>datatype</i>	::=	<i>BuiltIn</i> () <i>SelfDefined</i> ()	
<i>name</i>	::=	<i>Name</i> (<i>str</i>)	
<i>val</i>	::=	<i>Num</i> (<i>str</i>) <i>Empty</i> ()	
<i>pos</i>	::=	<i>Pos</i> (<i>Num</i> (<i>str</i>), <i>Num</i> (<i>str</i>)) <i>Empty</i> ()	
<i>size</i>	::=	<i>Num</i> (<i>str</i>) <i>Empty</i> ()	

1.3.1.3.3 Codebeispiel

In Code 1.9 sieht man den **Abstract-Syntax-Tree** des **PiocC-ANF Passes** für das aus Unterkapitel 1.6 weitergeführte Beispiel, indem alle Statements und Ausdrücke in **A-Normalform** sind. Die **IfElse(exp, stmts, stmts)**-Knoten sind hier in **A-Normalform** gebracht worden, indem ihre **Komplexe Bedingung** vorgezogen wurde und das Ergebnis der **Komplexen Bedingung** einer **Location** zugewiesen ist und sie selbst das Ergebnis über den **Atomaren Ausdruck** `Stack(Num(str))` vom Stack lesen: `IfElse(Stack(Num(str)), stmts, stmts)`. **Funktionen** sind nur noch über die **Labels** von Blöcken zu erkennen, die den gleichen Bezeichner haben, wie die ursprüngliche Funktion und es lässt sich nur durch das **Nachverfolgen** der `GoTo(Name('label'))`-Knoten nachvollziehen, was ursprünglich zur Funktion gehörte.

```

1 File
2   Name './example_faculty_it.picoc_mon',
3   [
4     Block
5       Name 'faculty.6',
6       [
7         // Assign(Name('res'), Num('1'))
8         Exp(Num('1'))
9         Assign(Stackframe(Num('1')), Stack(Num('1')))
10        // While(Num('1'), [])
11        Exp(GoTo(Name('condition_check.5')))
12      ],
13    Block
14      Name 'condition_check.5',
15      [
16        // IfElse(Num('1'), [], [])
17        Exp(Num('1')),
18        IfElse
19          Stack
20            Num '1',
21            [
22              GoTo(Name('while_branch.4'))
23            ],
24            [
25              GoTo(Name('while_after.1'))
26            ]
27        ],
28      Block
29        Name 'while_branch.4',
30        [
31          // If(Atom(Name('n'), Eq('=='), Num('1')), [])
32          // IfElse(Atom(Name('n'), Eq('=='), Num('1')), [], [])
33          Exp(Stackframe(Num('0')))
34          Exp(Num('1'))
35          Exp(Atom(Stack(Num('2')), Eq('=='), Stack(Num('1')))),
36          IfElse
37            Stack
38              Num '1',
39              [
40                GoTo(Name('if.3'))
41              ],
42              [
43                GoTo(Name('if_else_after.2'))
44              ]
45        ],

```

```

46 Block
47   Name 'if.3',
48   [
49     // Return(Name('res'))
50     Exp(Stackframe(Num('1')))
51     Return(Stack(Num('1')))
52   ],
53 Block
54   Name 'if_else_after.2',
55   [
56     // Assign(Name('res'), BinOp(Name('n'), Mul('*'), Name('res')))
57     Exp(Stackframe(Num('0')))
58     Exp(Stackframe(Num('1')))
59     Exp(BinOp(Stack(Num('2')), Mul('*'), Stack(Num('1'))))
60     Assign(Stackframe(Num('1')), Stack(Num('1')))
61     // Assign(Name('n'), BinOp(Name('n'), Sub('-'), Num('1')))
62     Exp(Stackframe(Num('0')))
63     Exp(Num('1'))
64     Exp(BinOp(Stack(Num('2')), Sub('-'), Stack(Num('1'))))
65     Assign(Stackframe(Num('0')), Stack(Num('1')))
66     Exp(GoTo(Name('condition_check.5')))
67   ],
68 Block
69   Name 'while_after.1',
70   [
71     Return(Empty())
72   ],
73 Block
74   Name 'main.0',
75   [
76     StackMalloc(Num('2'))
77     Exp(Num('4'))
78     NewStackframe(Name('faculty'), GoTo(Name('addr@next_instr')))
79     Exp(GoTo(Name('faculty.6')))
80     RemoveStackframe()
81     Exp(ACC)
82     Exp(Call(Name('print'), [Stack(Num('1'))]))
83     Return(Empty())
84   ]
85 ]

```

Code 1.9: PicoC-ANF Pass für Codebespiel

1.3.1.4 RETI-Blocks Pass

1.3.1.4.1 Aufgabe

Die Aufgabe des **RETI-Blocks Passes** ist es die **Statements** in der **Blöcken**, die durch **PicoC-Knoten** im **Abstrakter Syntaxbaum** der Sprache L_{PicoC_ANF} dargestellt sind durch ihren entsprechenden **RETI-Knoten** zu ersetzen.

1.3.1.4.2 Abstrakte Syntax

Die **Abstrakte Syntax** der Sprache L_{RETI_Blocks} in Grammatik 1.3.4 ist verglichen mit der **Abstrakten Syntax** der Sprache L_{PicoC_ANF} in Grammatik 1.3.3 stark verändert, denn der Großteil der **PicoC-Knoten** wird in diesem Pass durch entsprechende **RETI-Knoten** ersetzt. Die einzigen verbleibenden **PicoC-Knoten**

sind $\text{Exp}(\text{GoTo}(\text{str}))$, $\text{Block}(\text{Name}(\text{str}), \langle \text{instr} \rangle^*)$ und $\text{File}(\text{Name}(\text{str}), \langle \text{block} \rangle^*)$, da das gesamte Konzept mit den **Blöcken** erst im **RETI-Pass** in Unterkapitel 1.3.8 aufgelöst wird.

reg	$::=$	$\text{ACC}() \mid \text{IN1}() \mid \text{IN2}() \mid \text{PC}() \mid \text{SP}() \mid \text{BAF}()$	L_RETI
		$\mid \text{CS}() \mid \text{DS}()$	
arg	$::=$	$\text{Reg}(\langle \text{reg} \rangle) \mid \text{Num}(\text{str})$	
rel	$::=$	$\text{Eq}() \mid \text{NEq}() \mid \text{Lt}() \mid \text{LtE}() \mid \text{Gt}() \mid \text{GtE}()$	
		$\mid \text{Always}() \mid \text{NOp}()$	
op	$::=$	$\text{Add}() \mid \text{Addi}() \mid \text{Sub}() \mid \text{Subi}() \mid \text{Mult}() \mid \text{Multi}()$	
		$\mid \text{Div}() \mid \text{Divi}() \mid \text{Mod}() \mid \text{Modi}() \mid \text{Oplus}() \mid \text{Oplusi}()$	
		$\mid \text{Or}() \mid \text{Ori}() \mid \text{And}() \mid \text{Andi}()$	
		$\mid \text{Load}() \mid \text{Loadin}() \mid \text{Loadi}() \mid \text{Store}() \mid \text{Storein}() \mid \text{Move}()$	
instr	$::=$	$\text{Instr}(\langle \text{op} \rangle, \langle \text{arg} \rangle^+) \mid \text{Jump}(\langle \text{rel} \rangle, \text{Num}(\text{str})) \mid \text{Int}(\text{Num}(\text{str}))$	
		$\mid \text{RTI}() \mid \text{Call}(\text{Name}(\text{'print'}), \langle \text{reg} \rangle) \mid \text{Call}(\text{Name}(\text{'input'}), \langle \text{reg} \rangle)$	
		$\mid \text{SingleLineComment}(\text{str}, \text{str})$	
		$\mid \text{Instr}(\text{Loadi}(), [\text{Reg}(\text{Acc}()), \text{GoTo}(\text{Name}(\text{str}))]) \mid \text{Jump}(\text{Eq}(), \text{GoTo}(\text{Name}(\text{str})))$	
instr	$::=$	$\text{Exp}(\text{GoTo}(\text{str}))$	L_PicoC
block	$::=$	$\text{Block}(\text{Name}(\text{str}), \langle \text{instr} \rangle^*)$	
file	$::=$	$\text{File}(\text{Name}(\text{str}), \langle \text{block} \rangle^*)$	

Grammatik 1.3.4: Abstrakte Syntax der Sprache L_{RETI_Blocks}

1.3.1.4.3 Codebeispiel

In Code 1.10 sieht man den **Abstract-Syntax-Tree** des **RETI-Blocks Passes** für das aus Unterkapitel 1.6 weitergeführte Beispiel, indem die **Statements**, die durch entsprechende **PicoC-Knoten** im **Abstrakt Syntax Tree** der Sprache L_{PicoC_ANF} in Grammatik 1.3.3 repräsentiert waren nun durch ihre entsprechenden **RETI-Knoten** ersetzt werden.

```

1 File
2   Name './example_faculty_it.reti_blocks',
3   [
4     Block
5       Name 'faculty.6',
6       [
7         # // Assign(Name('res'), Num('1'))
8         # Exp(Num('1'))
9         SUBI SP 1;
10        LOADI ACC 1;
11        STOREIN SP ACC 1;
12        # Assign(Stackframe(Num('1')), Stack(Num('1')))
13        LOADIN SP ACC 1;
14        STOREIN BAF ACC -3;
15        ADDI SP 1;
16        # // While(Num('1'), [])
17        # Exp(GoTo(Name('condition_check.5')))
18        Exp(GoTo(Name('condition_check.5')))
19      ],
20    Block
21      Name 'condition_check.5',
22      [
23        # // IfElse(Num('1'), [], [])
24        # Exp(Num('1'))

```

```

25     SUBI SP 1;
26     LOADI ACC 1;
27     STOREIN SP ACC 1;
28     # IfElse(Stack(Num('1')), [], [])
29     LOADIN SP ACC 1;
30     ADDI SP 1;
31     JUMP== GoTo(Name('while_after.1'));
32     Exp(GoTo(Name('while_branch.4')))
33 ],
34 Block
35     Name 'while_branch.4',
36     [
37         # // If(Atom(Name('n'), Eq('=='), Num('1')), [])
38         # // IfElse(Atom(Name('n'), Eq('=='), Num('1')), [], [])
39         # Exp(Stackframe(Num('0')))
40         SUBI SP 1;
41         LOADIN BAF ACC -2;
42         STOREIN SP ACC 1;
43         # Exp(Num('1'))
44         SUBI SP 1;
45         LOADI ACC 1;
46         STOREIN SP ACC 1;
47         LOADIN SP ACC 2;
48         LOADIN SP IN2 1;
49         SUB ACC IN2;
50         JUMP== 3;
51         LOADI ACC 0;
52         JUMP 2;
53         LOADI ACC 1;
54         STOREIN SP ACC 2;
55         ADDI SP 1;
56         # IfElse(Stack(Num('1')), [], [])
57         LOADIN SP ACC 1;
58         ADDI SP 1;
59         JUMP== GoTo(Name('if_else_after.2'));
60         Exp(GoTo(Name('if.3')))
61     ],
62 Block
63     Name 'if.3',
64     [
65         # // Return(Name('res'))
66         # Exp(Stackframe(Num('1')))
67         SUBI SP 1;
68         LOADIN BAF ACC -3;
69         STOREIN SP ACC 1;
70         # Return(Stack(Num('1')))
71         LOADIN SP ACC 1;
72         ADDI SP 1;
73         LOADIN BAF PC -1;
74     ],
75 Block
76     Name 'if_else_after.2',
77     [
78         # // Assign(Name('res'), BinOp(Name('n'), Mul('*'), Name('res')))
79         # Exp(Stackframe(Num('0')))
80         SUBI SP 1;
81         LOADIN BAF ACC -2;

```

```

82     STOREIN SP ACC 1;
83     # Exp(Stackframe(Num('1')))
84     SUBI SP 1;
85     LOADIN BAF ACC -3;
86     STOREIN SP ACC 1;
87     # Exp(BinOp(Stack(Num('2')), Mul('*'), Stack(Num('1'))))
88     LOADIN SP ACC 2;
89     LOADIN SP IN2 1;
90     MULT ACC IN2;
91     STOREIN SP ACC 2;
92     ADDI SP 1;
93     # Assign(Stackframe(Num('1')), Stack(Num('1')))
94     LOADIN SP ACC 1;
95     STOREIN BAF ACC -3;
96     ADDI SP 1;
97     # // Assign(Name('n'), BinOp(Name('n'), Sub('-'), Num('1')))
98     # Exp(Stackframe(Num('0')))
99     SUBI SP 1;
100    LOADIN BAF ACC -2;
101    STOREIN SP ACC 1;
102    # Exp(Num('1'))
103    SUBI SP 1;
104    LOADI ACC 1;
105    STOREIN SP ACC 1;
106    # Exp(BinOp(Stack(Num('2')), Sub('-'), Stack(Num('1'))))
107    LOADIN SP ACC 2;
108    LOADIN SP IN2 1;
109    SUB ACC IN2;
110    STOREIN SP ACC 2;
111    ADDI SP 1;
112    # Assign(Stackframe(Num('0')), Stack(Num('1')))
113    LOADIN SP ACC 1;
114    STOREIN BAF ACC -2;
115    ADDI SP 1;
116    # Exp(GoTo(Name('condition_check.5')))
117    Exp(GoTo(Name('condition_check.5')))
118  ],
119  Block
120    Name 'while_after.1',
121    [
122      # Return(Empty())
123      LOADIN BAF PC -1;
124    ],
125  Block
126    Name 'main.0',
127    [
128      # StackMalloc(Num('2'))
129      SUBI SP 2;
130      # Exp(Num('4'))
131      SUBI SP 1;
132      LOADI ACC 4;
133      STOREIN SP ACC 1;
134      # NewStackframe(Name('faculty'), GoTo(Name('addr@next_instr')))
135      MOVE BAF ACC;
136      ADDI SP 3;
137      MOVE SP BAF;
138      SUBI SP 4;

```

```

139     STOREIN BAF ACC 0;
140     LOADI ACC GoTo(Name('addr@next_instr'));
141     ADD ACC CS;
142     STOREIN BAF ACC -1;
143     # Exp(GoTo(Name('faculty.6')))
144     Exp(GoTo(Name('faculty.6')))
145     # RemoveStackframe()
146     MOVE BAF IN1;
147     LOADIN IN1 BAF 0;
148     MOVE IN1 SP;
149     # Exp(ACC)
150     SUBI SP 1;
151     STOREIN SP ACC 1;
152     LOADIN SP ACC 1;
153     ADDI SP 1;
154     CALL PRINT ACC;
155     # Return(Empty())
156     LOADIN BAF PC -1;
157 ]
158 ]

```

Code 1.10: RETI-Blocks Pass für Codebeispiel

Wenn der **Abstrakter Syntaxbaum** ausgegeben wird, ist die Darstellung nicht ausschließlich in **Abstrakter Syntax**, da die **RETI-Knoten** aus bereits im Unterkapitel 1.2.5.6 vermitteltem Grund in **Konkreter Syntax** ausgegeben werden.

1.3.1.5 RETI-Patch Pass

1.3.1.5.1 Aufgabe

Die Aufgabe des **RETI-Patch Passes** ist das **Ausbessern** (engl. to patch) des **Abstrakter Syntaxbaums**, durch:

- das **Einfügen** eines **start.<nummer>-Blockes**, welcher ein **GoTo(Name('main'))** zur **main-Funktion** enthält, wenn in manchen Fällen die **main-Funktion** **nicht** die erste Funktion ist und daher am Anfang zur **main-Funktion** gesprungen werden muss.
- das **Entfernen** von **GoTo()**'s, deren Sprung nur **eine** Adresse weiterspringen würde.
- das **Voranstellen** von **RETI-Knoten**, die vor jeder **Division** **Instr(Div(), args)** prüfen, ob, nicht durch 0 geteilt wird.²⁵
- das Überprüfen darauf, ob bestimmte **Immediates** **Im(str)** in Befehlen, wie z.B. **Jump(rel, Im(str))**, **Instr(Loadin(), [reg, reg, Im(str)])**, **Instr(Loadi(), [reg, Im(str)])** usw. **kleiner** -2^{21} oder **größer** $2^{21} - 1$ sind. Im Fall dessen, dass es so ist, muss der **gewünschte Zahlenwert** durch **Bitshiften** und Anwenden von **bitweise Oder** berechnet werden. Im Fall, dessen, dass der **Immediate** allerdings **kleiner** $-(2^{31})$ oder **größer** $2^{31} - 1$ ist, wird eine Fehlermeldung **TooLargeLiteral** ausgegeben.

1.3.1.5.2 Abstrakte Syntax

²⁵Das fällt unter die Themenbereiche des **Bachelorprojekts** und wird daher **nicht** genauer erläutert.

Die **Abstrakte Syntax** der Sprache L_{RETI_Patch} in Grammatik 1.3.5 ist im Vergleich zur **Abstrakten Syntax** der Sprache L_{RETI_Blocks} in Grammatik 1.3.4 kaum verändert. Es muss nur ein Knoten `Exit()` hinzugefügt werden, der im Falle einer **Division durch 0** die Ausführung des Programs beendet.

reg	$::=$	$ACC() \mid IN1() \mid IN2() \mid PC() \mid SP() \mid BAF()$ $\mid CS() \mid DS()$	L_{RETI}
arg	$::=$	$Reg(\langle reg \rangle) \mid Num(str)$	
rel	$::=$	$Eq() \mid NEq() \mid Lt() \mid LtE() \mid Gt() \mid GtE()$ $\mid Always() \mid NOP()$	
op	$::=$	$Add() \mid Addi() \mid Sub() \mid Subi() \mid Mult() \mid Multi()$ $\mid Div() \mid Divi() \mid Mod() \mid Modi() \mid Oplus() \mid Oplusi()$ $\mid Or() \mid Ori() \mid And() \mid Andi()$ $\mid Load() \mid Loadin() \mid Loadi() \mid Store() \mid Storein() \mid Move()$	
$instr$	$::=$	$Instr(\langle op \rangle, \langle arg \rangle+) \mid Jump(\langle rel \rangle, Num(str)) \mid Int(Num(str))$ $\mid RTI() \mid Call(Name('print'), \langle reg \rangle) \mid Call(Name('input'), \langle reg \rangle)$ $\mid SingleLineComment(str, str)$ $\mid Instr(Loadi(), [Reg(Acc()), GoTo(Name(str))]) \mid Jump(Eq(), GoTo(Name(str)))$	
$instr$	$::=$	$Exp(GoTo(str)) \mid Exit(Num(str))$	L_{PicoC}
$block$	$::=$	$Block(Name(str), \langle instr \rangle^*)$	
$file$	$::=$	$File(Name(str), \langle block \rangle^*)$	

Grammatik 1.3.5: Abstrakte Syntax der Sprache L_{RETI_Patch}

1.3.1.5.3 Codebeispiel

In Code 1.11 sieht man den **Abstract-Syntax-Tree** des **Pioco-Patch Passes** für das aus Unterkapitel 1.6 weitergeführte Beispiel. Durch den **RETI-Patch Pass** wurde hier ein `start.<nummer>-Block`²⁶ eingesetzt, da die `main`-Funktion **nicht** die **erste** Funktion ist. Des Weiteren wurden durch diesen Pass einzelne `GoTo(Name(str))-Statements` entfernt²⁷, die nur einen Sprung um **eine** Position entsprochen hätten.

```

1 File
2   Name './example_faculty_it.reti_patch',
3   [
4     Block
5       Name 'start.7',
6       [
7         # // Exp(GoTo(Name('main.0')))
8         Exp(GoTo(Name('main.0')))
9       ],
10    Block
11      Name 'faculty.6',
12      [
13        # // Assign(Name('res'), Num('1'))
14        # Exp(Num('1'))
15        SUBI SP 1;
16        LOADI ACC 1;
17        STOREIN SP ACC 1;
18        # Assign(Stackframe(Num('1')), Stack(Num('1')))
19        LOADIN SP ACC 1;
20        STOREIN BAF ACC -3;

```

²⁶Dieser **Block** wurde im Code 1.8 markiert.

²⁷Diese **entfernten** `GoTo(Name(str))`'s wurden ebenfalls im Code 1.8 markiert.

```

21     ADDI SP 1;
22     # // While(Num('1'), [])
23     # Exp(GoTo(Name('condition_check.5')))
24     # // not included Exp(GoTo(Name('condition_check.5')))
25 ],
26 Block
27     Name 'condition_check.5',
28     [
29         # // IfElse(Num('1'), [], [])
30         # Exp(Num('1'))
31         SUBI SP 1;
32         LOADI ACC 1;
33         STOREIN SP ACC 1;
34         # IfElse(Stack(Num('1')), [], [])
35         LOADIN SP ACC 1;
36         ADDI SP 1;
37         JUMP== GoTo(Name('while_after.1'));
38         # // not included Exp(GoTo(Name('while_branch.4')))
39     ],
40 Block
41     Name 'while_branch.4',
42     [
43         # // If(Atom(Name('n'), Eq('='), Num('1')), [])
44         # // IfElse(Atom(Name('n'), Eq('='), Num('1')), [], [])
45         # Exp(Stackframe(Num('0')))
46         SUBI SP 1;
47         LOADIN BAF ACC -2;
48         STOREIN SP ACC 1;
49         # Exp(Num('1'))
50         SUBI SP 1;
51         LOADI ACC 1;
52         STOREIN SP ACC 1;
53         LOADIN SP ACC 2;
54         LOADIN SP IN2 1;
55         SUB ACC IN2;
56         JUMP== 3;
57         LOADI ACC 0;
58         JUMP 2;
59         LOADI ACC 1;
60         STOREIN SP ACC 2;
61         ADDI SP 1;
62         # IfElse(Stack(Num('1')), [], [])
63         LOADIN SP ACC 1;
64         ADDI SP 1;
65         JUMP== GoTo(Name('if_else_after.2'));
66         # // not included Exp(GoTo(Name('if.3')))
67     ],
68 Block
69     Name 'if.3',
70     [
71         # // Return(Name('res'))
72         # Exp(Stackframe(Num('1')))
73         SUBI SP 1;
74         LOADIN BAF ACC -3;
75         STOREIN SP ACC 1;
76         # Return(Stack(Num('1')))
77         LOADIN SP ACC 1;

```

```

78     ADDI SP 1;
79     LOADIN BAF PC -1;
80 ],
81 Block
82     Name 'if_else_after.2',
83     [
84         # // Assign(Name('res'), BinOp(Name('n'), Mul('*'), Name('res')))
85         # Exp(Stackframe(Num('0')))
86         SUBI SP 1;
87         LOADIN BAF ACC -2;
88         STOREIN SP ACC 1;
89         # Exp(Stackframe(Num('1')))
90         SUBI SP 1;
91         LOADIN BAF ACC -3;
92         STOREIN SP ACC 1;
93         # Exp(BinOp(Stack(Num('2')), Mul('*'), Stack(Num('1'))))
94         LOADIN SP ACC 2;
95         LOADIN SP IN2 1;
96         MULT ACC IN2;
97         STOREIN SP ACC 2;
98         ADDI SP 1;
99         # Assign(Stackframe(Num('1')), Stack(Num('1')))
100        LOADIN SP ACC 1;
101        STOREIN BAF ACC -3;
102        ADDI SP 1;
103        # // Assign(Name('n'), BinOp(Name('n'), Sub('-'), Num('1')))
104        # Exp(Stackframe(Num('0')))
105        SUBI SP 1;
106        LOADIN BAF ACC -2;
107        STOREIN SP ACC 1;
108        # Exp(Num('1'))
109        SUBI SP 1;
110        LOADI ACC 1;
111        STOREIN SP ACC 1;
112        # Exp(BinOp(Stack(Num('2')), Sub('-'), Stack(Num('1'))))
113        LOADIN SP ACC 2;
114        LOADIN SP IN2 1;
115        SUB ACC IN2;
116        STOREIN SP ACC 2;
117        ADDI SP 1;
118        # Assign(Stackframe(Num('0')), Stack(Num('1')))
119        LOADIN SP ACC 1;
120        STOREIN BAF ACC -2;
121        ADDI SP 1;
122        # Exp(GoTo(Name('condition_check.5')))
123        Exp(GoTo(Name('condition_check.5')))
124    ],
125 Block
126     Name 'while_after.1',
127     [
128         # Return(Empty())
129         LOADIN BAF PC -1;
130     ],
131 Block
132     Name 'main.0',
133     [
134         # StackMalloc(Num('2'))

```

```

135     SUBI SP 2;
136     # Exp(Num('4'))
137     SUBI SP 1;
138     LOADI ACC 4;
139     STOREIN SP ACC 1;
140     # NewStackframe(Name('faculty'), GoTo(Name('addr@next_instr')))
141     MOVE BAF ACC;
142     ADDI SP 3;
143     MOVE SP BAF;
144     SUBI SP 4;
145     STOREIN BAF ACC 0;
146     LOADI ACC GoTo(Name('addr@next_instr'));
147     ADD ACC CS;
148     STOREIN BAF ACC -1;
149     # Exp(GoTo(Name('faculty.6')))
150     Exp(GoTo(Name('faculty.6')))
151     # RemoveStackframe()
152     MOVE BAF IN1;
153     LOADIN IN1 BAF 0;
154     MOVE IN1 SP;
155     # Exp(ACC)
156     SUBI SP 1;
157     STOREIN SP ACC 1;
158     LOADIN SP ACC 1;
159     ADDI SP 1;
160     CALL PRINT ACC;
161     # Return(Empty())
162     LOADIN BAF PC -1;
163 ]
164 ]

```

Code 1.11: RETI-Patch Pass für Codebeispiel

1.3.1.6 RETI Pass

1.3.1.6.1 Aufgabe

Die Aufgabe des **RETI-Patch Passes** ist es die `GoTo(Name(str))`-Knoten in den den Knoten `Instr(Loadi(), [reg, GoTo(Name(str))])`, `Jump(Eq(), GoTo(Name(str)))` und `Exp(GoTo(Name(str)))` durch eine entsprechende **Adresse** zu ersetzen, die entsprechende **Distanz** oder einen entsprechenden **Sprungbefehl** mit passender **Distanz** `Jump(Always(), Im(str(distance)))`. Die **Distanz**- und **Adressberechnung** wird in Unterkapitel ?? genauer mit **Formeln** erklärt.

1.3.1.6.2 Konkrete und Abstrakte Syntax

Die **Abstrakte Syntax** der Sprache L_{RETI} in Grammatik 1.3.8 hat im Vergleich zur **Abstrakten Syntax** der Sprache $L_{RETI.Patch}$ in Grammatik 1.3.5 nur noch ausschließlich **RETI-Knoten**. Alle **RETI-Knoten** stehen nun einem `Program(Name(str), instr)`-Knoten.

Ausgegeben wird der finale **Maschinencode** allerdings in **Konkreter Syntax**, die sich aus den Grammatiken 1.3.6 und 1.3.7 für jeweils die **Lexikalische** und **Syntaktische Analyse** zusammensetzt. Der Grund, warum die **Konkrete Syntax** der Sprache L_{RETI} auch nochmal in einen Teil für die **Lexikalische** und **Syntaktische Analyse** unterteilt ist, hat den Grund, dass für die Bachelorarbeit zum **Testen** des **PicoC-Compilers** ein **RETI-Interpreter** implementiert wurde, der den RETI-Code **lexen** und **parsen** muss, um ihn später **interpretieren** zu können.

<i>dig_no_0</i>	::=	"1" "2" "3" "4" "5" "6"	<i>L_Program</i>
		"7" "8" "9"	
<i>dig_with_0</i>	::=	"0" <i>dig_no_0</i>	
<i>num</i>	::=	"0" <i>dig_no_0</i> <i>dig_with_0</i> * "-" <i>dig_no_0</i> *	
<i>letter</i>	::=	"a"..."Z"	
<i>name</i>	::=	<i>letter</i> (<i>letter</i> <i>dig_with_0</i> _)*	
<i>reg</i>	::=	"ACC" "IN1" "IN2" "PC" "SP"	
		"BAF" "CS" "DS"	
<i>arg</i>	::=	<i>reg</i> <i>num</i>	
<i>rel</i>	::=	"==" "!=" "<" "<=" ">"	
		">=" "_NOP"	

Grammatik 1.3.6: Konkrete Syntax der Sprache *L_{RETI}* für die Lexikalische Analyse in EBNF

<i>instr</i>	::=	"ADD" <i>reg</i> <i>arg</i> "ADDI" <i>reg</i> <i>num</i> "SUB" <i>reg</i> <i>arg</i>	<i>L_Program</i>
		"SUBI" <i>reg</i> <i>num</i> "MULT" <i>reg</i> <i>arg</i> "MULTI" <i>reg</i> <i>num</i>	
		"DIV" <i>reg</i> <i>arg</i> "DIVI" <i>reg</i> <i>num</i> "MOD" <i>reg</i> <i>arg</i>	
		"MODI" <i>reg</i> <i>num</i> "OPLUS" <i>reg</i> <i>arg</i> "OPLUSI" <i>reg</i> <i>num</i>	
		"OR" <i>reg</i> <i>arg</i> "ORI" <i>reg</i> <i>num</i>	
		"AND" <i>reg</i> <i>arg</i> "ANDI" <i>reg</i> <i>num</i>	
		"LOAD" <i>reg</i> <i>num</i> "LOADIN" <i>arg</i> <i>arg</i> <i>num</i>	
		"LOADI" <i>reg</i> <i>num</i>	
		"STORE" <i>reg</i> <i>num</i> "STOREIN" <i>arg</i> <i>argnum</i>	
		"MOVE" <i>reg</i> <i>reg</i>	
		"JUMP" <i>rel</i> <i>num</i> <i>INT</i> <i>num</i> <i>RTI</i>	
		"CALL" "INPUT" <i>reg</i> "CALL" "PRINT" <i>reg</i>	
<i>program</i>	::=	<i>name</i> (<i>instr</i> ";")*	

Grammatik 1.3.7: Konkrete Syntax der Sprache *L_{RETI}* für die Syntaktische Analyse in EBNF

<i>reg</i>	::=	<i>ACC</i> () <i>IN1</i> () <i>IN2</i> () <i>PC</i> () <i>SP</i> () <i>BAF</i> ()	<i>L_RETI</i>
		<i>CS</i> () <i>DS</i> ()	
<i>arg</i>	::=	<i>Reg</i> (<i><reg></i>) <i>Num</i> (<i>str</i>)	
<i>rel</i>	::=	<i>Eq</i> () <i>NEq</i> () <i>Lt</i> () <i>LtE</i> () <i>Gt</i> () <i>GtE</i> ()	
		<i>Always</i> () <i>NOp</i> ()	
<i>op</i>	::=	<i>Add</i> () <i>Addi</i> () <i>Sub</i> () <i>Subi</i> () <i>Mult</i> () <i>Multi</i> ()	
		<i>Div</i> () <i>Divi</i> () <i>Mod</i> () <i>Modi</i> () <i>Oplus</i> () <i>Oplusi</i> ()	
		<i>Or</i> () <i>Ori</i> () <i>And</i> () <i>Andi</i> ()	
		<i>Load</i> () <i>Loadin</i> () <i>Loadi</i> () <i>Store</i> () <i>Storein</i> () <i>Move</i> ()	
<i>instr</i>	::=	<i>Instr</i> (<i><op></i> , <i><arg></i> +) <i>Jump</i> (<i><rel></i> , <i>Num</i> (<i>str</i>)) <i>Int</i> (<i>Num</i> (<i>str</i>))	
		<i>RTI</i> () <i>Call</i> (<i>Name</i> ('print'), <i><reg></i>) <i>Call</i> (<i>Name</i> ('input'), <i><reg></i>)	
		<i>SingleLineComment</i> (<i>str</i> , <i>str</i>)	
		<i>Instr</i> (<i>Loadi</i> (), [<i>Reg</i> (<i>Acc</i> ()), <i>GoTo</i> (<i>Name</i> (<i>str</i>))]) <i>Jump</i> (<i>Eq</i> (), <i>GoTo</i> (<i>Name</i> (<i>str</i>)))	
<i>program</i>	::=	<i>Program</i> (<i>Name</i> (<i>str</i>), <i><instr></i> *)	
<i>instr</i>	::=	<i>Exp</i> (<i>GoTo</i> (<i>str</i>)) <i>Exit</i> (<i>Num</i> (<i>str</i>))	<i>L_PicoC</i>
<i>block</i>	::=	<i>Block</i> (<i>Name</i> (<i>str</i>), <i><instr></i> *)	
<i>file</i>	::=	<i>File</i> (<i>Name</i> (<i>str</i>), <i><block></i> *)	

Grammatik 1.3.8: Abstrakte Syntax der Sprache *L_{RETI}*

1.3.1.6.3 Codebeispiel

Nach dem **RETI-Pass** ist das Programm komplett in **RETI-Knoten** übersetzt, die allerdings in ihrer **Konkreten Syntax** ausgegeben werden, wie in Code 1.12 zu sehen ist. Es gibt **keine Blöcke** mehr und die **RETI-Befehle** in diesen Blöcken wurden **zusammengesetzt**, wie sie in den **Blöcken** angeordnet waren. Die letzten **Nicht-RETI-Befehle** oder **RETI-Befehle**, die **nicht** ausschließlich aus **RETI-Ausdrücken** bestehen²⁸, die sich in den **Blöcken** befunden haben, wurden durch **RETI-Befehle** ersetzt.

Der **Program(Name(str), instr)**-Knoten, indem alle **RETI-Knoten** stehen gibt alleinig die **RETI-Knoten**, die er beinhaltet aus und fügt ansonsten nichts hinzu, wodurch der **Abstrakter Syntaxbaum**, wenn er in eine Datei ausgegeben wird, direkt **RETI-Code** in **menschenlesbarer Repräsentation** erzeugt.

```

1 # // Exp(GoTo(Name('main.0'))))
2 JUMP 67;
3 # // Assign(Name('res'), Num('1'))
4 # Exp(Num('1'))
5 SUBI SP 1;
6 LOADI ACC 1;
7 STOREIN SP ACC 1;
8 # Assign(Stackframe(Num('1')), Stack(Num('1'))))
9 LOADIN SP ACC 1;
10 STOREIN BAF ACC -3;
11 ADDI SP 1;
12 # // While(Num('1'), [])
13 # Exp(GoTo(Name('condition_check.5'))))
14 # // not included Exp(GoTo(Name('condition_check.5'))))
15 # // IfElse(Num('1'), [], [])
16 # Exp(Num('1'))
17 SUBI SP 1;
18 LOADI ACC 1;
19 STOREIN SP ACC 1;
20 # IfElse(Stack(Num('1')), [], [])
21 LOADIN SP ACC 1;
22 ADDI SP 1;
23 JUMP== 54;
24 # // not included Exp(GoTo(Name('while_branch.4'))))
25 # // If(Atom(Name('n'), Eq('=='), Num('1')), [])
26 # // IfElse(Atom(Name('n'), Eq('=='), Num('1')), [], [])
27 # Exp(Stackframe(Num('0'))))
28 SUBI SP 1;
29 LOADIN BAF ACC -2;
30 STOREIN SP ACC 1;
31 # Exp(Num('1'))
32 SUBI SP 1;
33 LOADI ACC 1;
34 STOREIN SP ACC 1;
35 LOADIN SP ACC 2;
36 LOADIN SP IN2 1;
37 SUB ACC IN2;
38 JUMP== 3;
39 LOADI ACC 0;
40 JUMP 2;
41 LOADI ACC 1;
42 STOREIN SP ACC 2;

```

²⁸Wie z.B. `LOADI ACC GoTo(Name('addr@next_instr')), Exp(GoTo(Name('main.0')))` und `JUMP== GoTo(Name('if_else_after.2'))`.

```

43 ADDI SP 1;
44 # IfElse(Stack(Num('1')), [], [])
45 LOADIN SP ACC 1;
46 ADDI SP 1;
47 JUMP== 7;
48 # // not included Exp(GoTo(Name('if.3')))
49 # // Return(Name('res'))
50 # Exp(Stackframe(Num('1')))
51 SUBI SP 1;
52 LOADIN BAF ACC -3;
53 STOREIN SP ACC 1;
54 # Return(Stack(Num('1')))
55 LOADIN SP ACC 1;
56 ADDI SP 1;
57 LOADIN BAF PC -1;
58 # // Assign(Name('res'), BinOp(Name('n'), Mul('*'), Name('res')))
59 # Exp(Stackframe(Num('0')))
60 SUBI SP 1;
61 LOADIN BAF ACC -2;
62 STOREIN SP ACC 1;
63 # Exp(Stackframe(Num('1')))
64 SUBI SP 1;
65 LOADIN BAF ACC -3;
66 STOREIN SP ACC 1;
67 # Exp(BinOp(Stack(Num('2')), Mul('*'), Stack(Num('1'))))
68 LOADIN SP ACC 2;
69 LOADIN SP IN2 1;
70 MULT ACC IN2;
71 STOREIN SP ACC 2;
72 ADDI SP 1;
73 # Assign(Stackframe(Num('1')), Stack(Num('1')))
74 LOADIN SP ACC 1;
75 STOREIN BAF ACC -3;
76 ADDI SP 1;
77 # // Assign(Name('n'), BinOp(Name('n'), Sub('-'), Num('1')))
78 # Exp(Stackframe(Num('0')))
79 SUBI SP 1;
80 LOADIN BAF ACC -2;
81 STOREIN SP ACC 1;
82 # Exp(Num('1'))
83 SUBI SP 1;
84 LOADI ACC 1;
85 STOREIN SP ACC 1;
86 # Exp(BinOp(Stack(Num('2')), Sub('-'), Stack(Num('1'))))
87 LOADIN SP ACC 2;
88 LOADIN SP IN2 1;
89 SUB ACC IN2;
90 STOREIN SP ACC 2;
91 ADDI SP 1;
92 # Assign(Stackframe(Num('0')), Stack(Num('1')))
93 LOADIN SP ACC 1;
94 STOREIN BAF ACC -2;
95 ADDI SP 1;
96 # Exp(GoTo(Name('condition_check.5')))
97 JUMP -58;
98 # Return(Empty())
99 LOADIN BAF PC -1;

```

```
00 # StackMalloc(Num('2'))
01 SUBI SP 2;
02 # Exp(Num('4'))
03 SUBI SP 1;
04 LOADI ACC 4;
05 STOREIN SP ACC 1;
06 # NewStackframe(Name('faculty'), GoTo(Name('addr@next_instr'))
07 MOVE BAF ACC;
08 ADDI SP 3;
09 MOVE SP BAF;
10 SUBI SP 4;
11 STOREIN BAF ACC 0;
12 LOADI ACC 80;
13 ADD ACC CS;
14 STOREIN BAF ACC -1;
15 # Exp(GoTo(Name('faculty.6'))
16 JUMP -78;
17 # RemoveStackframe()
18 MOVE BAF IN1;
19 LOADIN IN1 BAF 0;
20 MOVE IN1 SP;
21 # Exp(ACC)
22 SUBI SP 1;
23 STOREIN SP ACC 1;
24 LOADIN SP ACC 1;
25 ADDI SP 1;
26 CALL PRINT ACC;
27 # Return(Empty())
28 LOADIN BAF PC -1;
```

Code 1.12: *RETI Pass für Codebeispiel*

Literatur

Online

- *ANSI C grammar (Lex)*. URL: <https://www.lysator.liu.se/c/ANSI-C-grammar-1.html> (besucht am 29.07.2022).
- *ANSI C grammar (Yacc)*. URL: <http://www.quot.com/c/ANSI-C-grammar-y.html> (besucht am 29.07.2022).
- *ANTLR*. URL: <https://www.antlr.org/> (besucht am 31.07.2022).
- *C Operator Precedence - cppreference.com*. URL: https://en.cppreference.com/w/c/language/operator_precedence (besucht am 27.04.2022).
- *clang: C++ Compiler*. URL: <http://clang.org/> (besucht am 29.07.2022).
- *Clockwise/Spiral Rule*. URL: <https://c-faq.com/decl/spiral.anderson.html> (besucht am 29.07.2022).
- *GCC, the GNU Compiler Collection - GNU Project*. URL: <https://gcc.gnu.org/> (besucht am 13.07.2022).
- *Grammar Reference — Lark documentation*. URL: <https://lark-parser.readthedocs.io/en/latest/grammar.html> (besucht am 31.07.2022).
- *Grammar: The language of languages (BNF, EBNF, ABNF and more)*. URL: <https://matt.might.net/articles/grammars-bnf-ebnf/> (besucht am 30.07.2022).
- *Welcome to Lark's documentation! — Lark documentation*. URL: <https://lark-parser.readthedocs.io/en/latest/> (besucht am 31.07.2022).

Bücher

- G. Siek, Jeremy. *Course Webpage for Compilers (P423, P523, E313, and E513)*. 28. Jan. 2022. URL: <https://iucompilercourse.github.io/IU-Fall-2021/> (besucht am 28.01.2022).

Artikel

- Earley, Jay. „An efficient context-free parsing“. In: 13 (1968). URL: <https://web.archive.org/web/20040708052627/http://www-2.cs.cmu.edu/afs/cs.cmu.edu/project/cmt-55/lti/Courses/711/Class-notes/p94-earley.pdf> (besucht am 10.08.2022).

Vorlesungen

- Nebel, Prof. Dr. Bernhard. „Theoretische Informatik“. Vorlesung. Vorlesung. Universität Freiburg, 2020. URL: http://gki.informatik.uni-freiburg.de/teaching/ss20/info3/index_de.html (besucht am 09.07.2022).
- Scholl, Prof. Dr. Christoph. „Betriebssysteme“. Vorlesung. Vorlesung. Universität Freiburg, 2020. URL: https://abs.informatik.uni-freiburg.de/src/teach_main.php?id=157 (besucht am 09.07.2022).
- Thiemann, Prof. Dr. Peter. „Compilerbau“. Vorlesung. Vorlesung. Universität Freiburg, 2021. URL: <http://proglang.informatik.uni-freiburg.de/teaching/compilerbau/2021ws/> (besucht am 09.07.2022).

Sonstige Quellen

- *Lark - a parsing toolkit for Python*. 26. Apr. 2022. URL: <https://github.com/lark-parser/lark> (besucht am 28.04.2022).
- Shinan, Erez. *lark: a modern parsing library*. Version 1.1.2. URL: <https://github.com/lark-parser/lark> (besucht am 31.07.2022).
- *Syntax*. In: *Wiktionary*. Page Version ID: 9196998. 7. Juni 2022. URL: <https://de.wiktionary.org/w/index.php?title=Syntax&oldid=9196998> (besucht am 31.07.2022).