IMPLEMENTATION OF A COMPILER FOR THE FUNCTIONAL
PROGRAMMING LANGUAGE PHI(U) NAVAL POSTGRADUATE SCHOOL
MONTEREY CA  E J COLE ET AL. JUN 87
Thesis

Implementation of a Compiler for the Functional Programming Language SHL - 1

by

Eugene J. Cole

and

Joseph E. Donnell II

June 1987

Thesis Advisor: Daniel Davis

Approved for public release; distribution is unlimited.
TITLe: FUNCTIONAL LEXICAL SYNTAX: APPLICATION TO AN ALGOL 60 COMPILER

This paper describes the design and implementation of a prototype compiler for a functional language, F:4J. The design is that of a lexical syntax-independent lexical and syntactic analyzer and code generator. The back-end of the compiler implements a machine-independent lexical and syntactic analyzer and code generator. The back-end implements a machine-independent lexical and syntactic analyzer and code generator. The front-end of the compiler implements a machine-independent lexical and syntactic analyzer and code generator.

The implementation of this prototype is a prototype, and it does not contain all of the features of a full implementation. The basic features, such as functions and data definitions are implemented, as well as integer, rational, and boolean types. However, the necessary features are present and the design is mature enough to allow expansion into a full implementation.

SUBJECT TERMS: Functional Language; Application to Compiler Design.
Approved for public release; distribution is unlimited.

Implementation of a Compiler for the
Functional Programming Language PHI — Φ

by

Eugene J. Cole
Major, United States Marine Corps
B. A., The Citadel, 1975

and

Joseph E. Connell II
Captain, United States Marine Corps
B. S., University of Missouri — Rolla, 1974

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN COMPUTER SCIENCE

from the

NAVAL POSTGRADUATE SCHOOL
June 1987

Authors:

Eugene J. Cole

Joseph E. Connell II

Approved by:

Daniel Davis, Thesis Advisor

Bruce J. MacLennan, Second Reader

Vincent Y. Lam, Chairman,
Department of Computer Science

Kneale T. Marshall,
Dean of Information and Policy Sciences
ABSTRACT

This thesis describes the design and implement of a prototype compiler for the functional programming language PHI. The design is highly modularized and the authors think this should facilitate the understanding of both concept and implementation. The front-end of the compiler implements machine independent lexical and syntactic analyzers; top-down parsing techniques are employed. The back-end implements a machine dependent one-pass semantic analyzer and code generator.

Since this implementation is a prototype, it does not possess all of the qualities desirable in a full implementation. The basic constructs of PHI: functions and data definitions are implemented, as well as the integer, natural number, and boolean types. However, the necessary hooks are present and the design is mature enough to allow expanding the prototype to a full implementation.
# TABLE OF CONTENTS

I. INTRODUCTION ....................................................................................... 6  
   A. BACKGROUND — GENERAL ................................................................. 6  
   B. BACKGROUND — THESIS ...................................................................... 7  
   C. BACKGROUND — FUNCTIONAL PROGRAMMING .................................... 8  
      1. Problems with Conventional Languages ........................................... 9  
      2. Functional Languages ..................................................................... 10  
   D. ASSUMPTIONS .................................................................................... 11  
   E. CONSTRAINTS ..................................................................................... 12  

II. FRONT-END OF THE COMPILER ........................................................... 13  
   A. LEXICAL ANALYSIS — THE SCANNER ............................................ 13  
   B. SYNTACTIC ANALYSIS — THE PARSER .......................................... 18  
   C. ERROR HANDLING ........................................................................... 23  

III. BACK-END OF THE COMPILER ........................................................... 25  
   A. OVERVIEW ......................................................................................... 25  
   B. RUN-TIME ORGANIZATION .............................................................. 25  
   C. SEMANTIC CHECKING AND CODE GENERATION .............................. 31  
   D. OPTIMIZATION .................................................................................. 40  

IV. RESULTS & CONCLUSIONS ................................................................. 42  
   A. RESULTS ............................................................................................ 42  
   B. CONCLUSIONS .................................................................................. 43  

V. FURTHER RESEARCH ........................................................................... 44  

LIST OF REFERENCES .............................................................................. 46
I. INTRODUCTION

A. BACKGROUND — GENERAL

In its attempt to provide students with a well rounded background to the field of computer science, the computer science department at the Naval Postgraduate School offers courses covering recent developments in programming languages. One of the courses deals specifically with the methodology of functional, also known as applicative, programming. Both the theory and the practice of functional programming are covered, concentrating more on the practice than the theory. In order to fully appreciate the nuances of functional programming it would be desirable to provide the students with a functional programming environment. This would provide a first hand look at the fundamental difference in methodologies when programming in functional languages as opposed to programming in traditional imperative languages.

Of the languages currently supported in the department; LISP, on the UNIX environment, comes the closest to meeting this requirement. Although LISP is considered a functional language by some, its many extensions and modifications actually brings it into the world of imperative programming. It is not a pure functional programming language.

There are several additional problems associated with using LISP to teach techniques of functional programming. Modern LISP dialects do not support all aspects of functional programming. Most notably they lack the ability to define higher-order functions. Dynamic scoping and the semantics of the language make it a pedagogical nightmare to teach.[Ref. 1:p. 9-1] The goal of teaching functional programming would rapidly be overtaken by the necessity of explaining the idiosyncrasies of LISP. In an 11 week

---

1 UNIX is a trademark of Bell Laboratories.
quarter, time devoted to LISP would significantly detract from instruction of functional programming.

Recognizing the shortcomings of LISP, a pure functional language, PHI was developed by Dr. B. J. MacLennan for use in this course of instruction. The syntax of PHI closely follows that of standard mathematical notation. This means students should have little difficulty in learning how to write legitimate PHI statements. Instruction can now concentrate on joining these statements to create functional programs. Hopefully, this will lead to a greater understanding and appreciation of the methodology of functional programming.

B. BACKGROUND — THESIS

Creation of PHI solved the problem of finding a suitable language to use to demonstrate the methodology of functional programming. However, currently PHI programs are programs on paper only. There exists no programming environment for the PHI language. So it is impossible to machine execute PHI programs. This thesis attempts to remedy the above problem by providing the first component in a PHI programming environment — a prototype PHI compiler.

Conventional compiler construction techniques were chosen for this implementation for several reasons. By choosing conventional techniques, the authors were able to address the problems associated with utilizing conventional methods for implementing a compiler for a functional language. Additionally, realizing that both the language and system would change, the authors wanted a well documented and understood methodology. The cost of maintaining a system can be as much as three times the development cost [Ref. 2:p. 478]. Therefore, it was imperative to choose a methodology that supported a clean and structured design.

---

2 Specific problems and solutions are covered later in Chapters Two and Three

7
Following conventional methodologies, the authors separated the PHI compiler design into a front-end and a back-end. The overall general design of the PHI compiler is shown in Figure 1.1. The front-end, containing the scanner (lexical analyzer) and parser (syntactic analyzer) is essentially responsible for analysis of the external file containing the source program. The PHI compiler back-end couples semantic analysis with code generation to produce code suitable for execution on the target machine. The authors felt that a clear and distinct separation between parts would aid understanding of the system, simplify division of labor, and increase ease of development and maintenance. It should also result in greater flexibility for follow-on development in the PHI programming environment. As an example, the current front-end could be modified to support a PHI interpreter.

![Figure 1.1 General Design of the PHI Compiler](image)

**C. BACKGROUND — FUNCTIONAL PROGRAMMING**

Functional programming is a methodology in favor among academicians. Although applicative programming goes further back, it is generally agreed that, as a methodology, functional programming traces its roots to John Backus [Ref. 4:p. 404, Ref. 5:p. 65].

---

3Design and implementation of the front-end is discussed in Chapter Two.

4Design and implementation of the back-end is discussed in Chapter Three.
his acceptance speech for the 1977 ACM Turing Award, Backus criticized traditional programming languages and programming styles. He went on to propose a new methodology of programming that involved "the use of a fixed set of combining forms called functional forms." [Ref. 6:p. 619] This methodology is known today as functional programming.

1. Problems with Conventional Languages

Backus feels [Ref. 6:pp. 613–619] that the basic underlying problem with conventional languages is the existence of the assignment statement. The assignment statement plays a central role in conventional languages and breaks programming into two worlds. Backus calls the right-hand side of assignment statements, expressions, the first of these worlds. The second world is the world of statements, with the primary statement, of course, being the assignment statement.

Several problems are associated with assignment statements. First, they permit programs to be held hostage through access to their variables. Since variables are used to imitate the machine’s storage cells; assignment statements allow, even encourage, state changes to take place. This access, either direct or indirect, permits such problems as side effects, unintentional state changes, and aliasing to arise. It then becomes difficult to reason about the correctness of these programs, so proving simple programs correct is an arduous task and proving complex programs correct is virtually impossible. Additionally, by permitting the value of variables to be changed, the assignment statement makes temporal order of execution of statements critical. For example, the following two pieces of code produce dramatically different results depending on which statement inside the for loop is executed first.

```plaintext
for (i = 0; i != some_value; ++i)
    if (i % 2 == 0) continue;
    DoSomething(i);

for (i = 0; i != some_value; ++i)
    if (i % 2 == 0) continue;
    DoSomething(i);
```

These problems interact so that it becomes extremely difficult to create new programs out of old ones. [Ref. 6:pp. 613 - 619, Ref. 1:pp. 1-2 - 1-20]

Another problem associated with assignment statements is that each produces only a one-word result. In effect, they force programmers to think in a word-at-a-time manner. For example, to apply a function to an entire array of values, the programmer must access each value individually. Not only is this wasteful of computer assets, but it results in what Backus refers to as the "von Neumann bottleneck" of conventional programming languages. [Ref. 6:pp. 613 - 619]

2. Functional Languages

Backus proposes the methodology of functional programming as the solution to these problems. Functional languages have removed variables and the assignment statement from their syntax so that their basic building block becomes the function. It is through "the use of a fixed set of combining forms...plus simple definitions" [Ref. 6:p. 619] that the programmer is able to build new functions from existing functions. It thus becomes possible to form a new program by combining two or more existing programs or functions together.

The absence of assignment statements and variables removes the problems plaguing conventional languages caused by side effects, etc. because the program now operates exclusively in the world of expressions. This permits the programmer to maintain a clear conceptual view of the program. It is easier to understand and reason about the task the program is to perform [Ref. 5:pp. 65 - 69]. It now becomes not only possible, but practical to prove programs correct [Ref.6:pp. 624 - 625].

Another direct benefit stemming from the absence of side effects is order. The values of expressions are no longer dependent on the order in which they are evaluated. Therefore, functional languages provide a natural means of performing parallel computations [Ref. 7:p. 35]. Functional languages and the associated methodology of
functional programming may very well provide the key to programming the massively parallel computers entering service nowadays. All of the above benefits have applicability to ongoing research in the SDI program.

The authors feel that functional programming can best be summarized by the following thought — assignment statements are to functional programming what GOTO statements are to structured programming.

D. ASSUMPTIONS

An IBM\textsuperscript{5} personal computer/IBM compatible personal computer was chosen as the target machine for this implementation. The authors felt that the nature of the language and its intended use were better suited for the PC/personal work station environment as opposed to a mini– or main–frame time shared environment. The PC environment should provide greater flexibility and freedom when implementing follow–on tools for the PHI programming language. Also, future compiler improvements will not have to be concerned with extraneous interfaces to another system. Working with a PC environment eliminates the need to take into account the effects the PHI environment will have on another user of the system. The implementor is able to work with a system that remains constant — a known quantity.

The assumed target machine configuration is based on the equipment available in the Naval Postgraduate School’s computer science microcomputer lab. Each machine is configured with 640K bytes of RAM, one (most have two) 20M byte hard disk drive, one 1.2M byte 5 inch floppy disk drive, and the 8087 math co–processor: each currently operates under the MS–DOS\textsuperscript{6} 3.x operating system. These machines are readily available to all computer science students at the Naval Postgraduate School, and many students own

\textsuperscript{5}IBM is a registered trademark of Internal Business Machines Corporation.

\textsuperscript{6}MS–DOS is a registered trademark of Microsoft Corporation.
personal computers with similar configurations. It is not necessary to utilize a hard disk when executing the PHI compiler.

E. CONSTRAINTS

As is the case with most implementation theses, time was probably the biggest constraint facing the authors. This involved making certain trade-offs; e.g. should the major effort be directed towards a full implementation of PHI while concentrating on a particular component of the compiler, or should the major effort be directed towards a full implementation of the compiler while concentrating on a subset of the PHI language? The authors felt that the greatest benefit could be gained by implementing a complete compiler. Having to actually face the issues and problems associated with designing, implementing, and interfacing a full compiler implementation would be much different than just reading about them in a text. As a result, this thesis implements only a subset\textsuperscript{7} of PHI.

Since PHI is an experimental language it is still undergoing changes and revisions. Trying to modify and update the compiler design with each version proved to be an impossibility. The authors were forced to freeze the design based on the language as it stood on 07 January 1987. Any follow-on work will need to update the front-end and back-end of the compiler to meet the requirements of these new versions of PHI. A description of the grammar as implemented and a description of the latest version of the grammar may be found in the Appendixes.

\textsuperscript{7}This subset is discussed in the individual chapters on the front-end and back-end.
II. FRONT-END OF THE COMPILER

The authors separated the design of the PHI compiler into two modules, a front-end and a back-end. These modules were then further subdivided to produce the general layout of Figure 1.1. The authors believe this modularization simplifies the design and will aid in understanding the system, thus decreasing future maintenance problems.

The front-end of the PHI compiler is comprised of the scanner (lexical analyzer), the parser (syntactic analyzer), and their associated error recovery routines. Two possible interactions between the lexical and syntactic analyzers were considered. The first incorporates the scanner into the parser, and tokens are produced by the scanner only upon request of the syntactic analyzer. Thus, this system acts like a pipeline. An alternate method is to allow the scanner to tokenize the entire source program, store the tokens in some data structure, and pass this structure to the parser. [Ref. 3:p. 10]

For the prototype implementation of a PHI compiler, the authors based the design on the first interaction. Although the second method is conceptually very easy to understand, the authors think the current implementation is clean and will readily lend itself to future enhancements. Any input alphabet peculiarities are restricted to the lexical analyzer, and this independence should provide benefits for the next student(s) who work on the PHI programming environment.

A. LEXICAL ANALYSIS — THE SCANNER

The PHI compiler reads a source file of ASCII text which is fed to the scanner for lexical analysis. The principle task of lexical analysis is to separate or divide the source program into tokens for use during syntactic analysis [Ref.8:p. 84, Ref. 9:p. 155]. This is accomplished in the PHI compiler through a character-by-character examination of the
user's source file. These characters are assembled/grouped into the individual tokens which represent terminal symbols of the PHI grammar. Examples of some of the terminal symbols are operators, identifiers, keywords, and constants. A complete listing of the PHI tokens may be found in the header file for the scanner in Appendix E.

The primary advantage to tokenizing the source program is that the design of the syntactic analyzer needs to take into account only one type of data unit — the token [Ref. 3:p. 7]. This simplifies the design of the parser because provisions do not have to be made for handling white space and comments. The scanner has already removed them. Also, removing white space and comments and utilizing a fixed-length representation for the tokens saves space. Once tokenization is complete, the source program can be discarded and the compacted tokenized file can be utilized for further analysis.

In order to correctly tokenize the source file there must be some discrete means available to accurately represent each token. There are several ways of describing tokens. One means available is to use a regular grammar. In this method "generative rules are given for producing the desired tokens" [Ref. 3:p. 142]. An equivalent, but different, method is to use finite-state acceptors, FSAs, to recognize tokens. The authors found it easier to visualize this as a recognitive vice generative problem. For this reason the various tokens were modeled using FSAs. An example of an unsigned number recognizer is shown in Figure 2.1. The interested reader is directed to Tremblay and Sorenson [Ref. 3:Chapter 4] for an excellent introduction to the practice of using FSAs to model tokens. The authors found that utilizing FSAs greatly simplified the design, coding, and debugging of the lexical analyzer — one picture was worth a hundred lines of code.

The ideal grammar would allow each token to be uniquely and unambiguously identified. Once the lexical analyzer started on the path of building a token, it would be able to continue until the end with no backtracking. Due to limitations with the standard
ASCII character set, the designer of PHI used multiple keystrokes, or characters, to represent various operators in the language. This resulted in compound token types. Also, as in other programming languages, PHI overloads certain operators, allowing them to do double duty by taking on different context-dependent meanings.

The problem of dealing with compound token types was easily handled through the use of a single lookahead character. For example, upon finding the character "-", the scanner looks ahead to the next character to see if it is "->" (→) or another "--" (→→). If the next character is neither of these two, it indicates that the token is just the simple token "-".

Distinguishing overloaded operators was solved by essentially ignoring it in the scanner! The authors took the position this is basically a syntax analyzer problem and there was no reason to complicate the scanner by handling it. The scanner just identifies a generic token type, e.g. SUB_, and lets the parser make the proper determination of its true meaning, e.g. SUB_ or NEG_.

There are several design decisions relating to the lexical analyzer worth noting. The authors, following the example of Pascal, C, and other languages, took the position that

---

8 Some examples of this are -> for →, == for = and <> for ±.

9 For example, + and - can serve as either an unary or binary arithmetic operator.
PHI’s keywords\textsuperscript{10} are reserved words and may not be redefined and used as identifiers. Alternate decisions would have been to distinguish keywords from identifiers based on context, as PL/I does, or to precede them by some special character, as ALGOL 60 and ALGOL 68 do [Ref. 3:p. 91]. PHI has a very small set of keywords, smaller than C’s, and the authors think that this decision makes life easier for the programmer by simplifying debugging of programs. It certainly made life easier for the authors.

PHI’s grammar makes no provisions for programmer comments. The authors originally implemented comments by requiring the programmer to explicitly indicate the beginning and end of each comment with a special character. After scanning the special character at the beginning of the comment, the lexical analyzer would ignore all following characters until the special character was once again found. Following conversations with PHI’s designer this implementation was changed. Comments are now implemented the same way they are in Ada\textsuperscript{11}: the comment terminator is the end–of–line character. Not only did this simplify the recognizer for comments, but it also completely removed the problem of runaway comments.

A name table is used to point to the names of all identifiers and constants. A symbol table was originally utilized but later discarded when the authors realized the syntax of PHI makes analyzing an abstract syntax tree easier than analyzing a flattened tree. The information normally associated with a symbol table is now held in the nodes of the tree. This permits just the first instance of each name to be placed into the name table. In other words, regardless of how many times and in how many scopes the identifier X is used, X appears only once in the name table. The token returned to the parser would indicate a

\textsuperscript{10}A complete listing of PHI keywords may be found in the header file for the scanner in Appendix E.

\textsuperscript{11}Ada is a trademark of the Ada Joint Programming Office, Department of Defense, United States Government.
token type of identifier and the parser would then know to dereference the pointer to find the string containing the actual name, X.

Because keywords are reserved, each potential identifier must first be compared against the possible keywords prior to being placed in the name table. The authors implemented a keyword table to simplify this process. Knuth [Ref. 10:pp. 406-410] has shown that a binary search is the most efficient way of searching an ordered table, using only comparisons. For this reason the keyword table is kept in alphabetical order. The lookup, which is at worst $O(\log n)$, is performed using a binary search of the keyword table.

In an attempt to improve the efficiency of the name table, the authors implemented it as a hash table. McKeeman [Ref. 11:pp. 253–301] experimented with six different length dependent hash functions. He found that the function producing the best results involved summing the internal representation of the first and last characters of the variable's name with its length shifted four places to the left. This was the function utilized by the authors. The possibility of collisions is reduced by choosing a prime number as the table size. However, since this only reduces, not eliminates, the possibility of two or more names hashing to the same value; the authors had to make provisions for handling collisions.

A variant of the chaining method of collision–resolution was chosen. In PHI's implementation, each of the name table slots/buckets holds a data structure that can contain both the name of the variable and a pointer to another structure of the same type. So each hashed value points to a linked list of names. This method offers the advantage of providing better performance than linear probing [Ref. 12:p. 89], is conceptually easy to visualize/work with, and also solves the problem of possibly overflowing the hash table. It does require slightly more memory to implement, but the authors determined that the benefits of this method far outweighed the slight increase in storage requirements. [Ref. 12:pp. 83–93]
B. SYNTACTIC ANALYSIS — THE PARSER

The purpose of the parser is twofold: 1) to determine if the program, as represented by the output from the scanner, is syntactically correct; 2) to impose a hierarchical structure on the token stream, fitting it into the abstract syntax tree which is the output of the parser [Ref. 8: pp. 7–8, Ref. 9: p. 7]. Traditionally, these tasks are done by either a top–down or bottom–up methodology [Ref. 8: p. 41]. Both methodologies use the tokens generated through lexical analysis.

The terminology top–down refers to the order in which the nodes of the parse tree are constructed. Top–down parsing starts from the root of the tree and proceeds downward towards the terminal symbols at the leaves. The parse tree is constructed from the top to the bottom by applying productions of the grammar to generate strings of terminals and nonterminals. On the other hand, bottom–up methodologies start from the terminal symbols at the leaves and proceed upwards to the root. The parse tree is constructed from the bottom to the top by applying reductions of the grammar to generate single nonterminals from strings of terminals and nonterminals. [Ref. 8: pp. 40–41, Ref. 9: pp. 134–136]

It is generally agreed that the popularity of top–down parsing techniques is "due to the fact that efficient parsers can be constructed more easily by hand". [Ref. 8: p. 41] The authors can attest to the fact that the concept of top–down parsing is very easy to grasp. When parsing PHI, it is natural to begin with the start symbol of the grammar, BLOCKBODY, and work forward from there to analyze the token stream. So, partially because of its efficiency, but primarily because of its ease of understanding and use, the authors chose the top–down methodology of recursive–descent parsing to design and implement the syntactic analyzer.

In recursive–descent parsers, separate procedures/functions are written to recognize each nonterminal of the grammar [Ref. 3: pp. 219–220]. This technique gets its distinctive name "because nonterminals can appear in the right–hand sides of each other's
productions, the procedures for recognizing nonterminals are recursive." [Ref.9:p. 150]

To state it more clearly, the function to recognize nonterminal 'A' could end up calling itself recursively if either 1) 'A' appears on the right-hand side of the production describing 'A' itself, or 2) 'A' appears on the right-hand side of the production describing another nonterminal 'B' and 'B' appears on the right-hand side of the production describing 'A'. Regardless of how one looks at the nature of the technique, one usually identifies a stack with recursion. What made this technique so easy to implement was that the authors were able to use C's underlying mechanism for handling recursive functions. The authors did not have to explicitly maintain a stack of symbols for each function call; instead, the information was implicit in the stack of activation records resulting from each function call.

Top-down parsing techniques, especially recursive descent, offer straightforward means of implementing a syntactic analyzer. However, these techniques are applicable only to a subset of the context-free grammars and it is essential that all left recursion be eliminated from the grammar [Ref. 3:p. 211]. In other words, there must not exist any productions describing nonterminal 'A' with 'A' appearing as the first element on the right-hand side of the production. Obviously, if this situation existed, it would be possible to present the parser with strings to parse that would cause it to enter "an infinite loop of production applications" [Ref. 3:p. 211], never to be heard from again. The PHI production QUALEXP = QUALEXP WHERE AUXDEFS is an example of this type of string. The parser would hang up looking for QUALEXP and would never leave this loop until the machine ran out of memory stacking activation records. In order to employ top-down parsing techniques with PHI the authors rewrote the PHI grammar to be right-recursive\(^2\). However, as shown below, even the new grammar does not lend itself to "pure" recursive descent parsing techniques.

\(^{12}\) The right recursive syntax of PHI may be found in Appendix D.
From the compiler writer's point of view the ideal grammar would allow the correct production rule to be applied in every step of the parsing process. Constructing the parse tree would then proceed in a completely deterministic manner. When this is not possible, there are two basic parser design methods for dealing with nondeterminism in the grammar [Ref. 9:pp. 151–152]. In the backtracking method, which is generally not applicable to recursive–descent techniques, the parser picks an arbitrary production and continues with the parse [Ref. 9:p. 151]. If the parse is successful it is assumed that the correct production was chosen. However, if an error is later discovered, the parser backtracks to the last choice, a new production is chosen, and the parser presses forward again. This process continues until either the parse is successful or the parser runs out of possible productions to choose from. The second method requires a modification to the grammar which results in a deterministic parser: the grammar is rewritten using a process called left factoring to avoid choices among nonterminals [Ref. 9:p. 151].

For the most part, the design of PHI is conducive to recursive descent parsing techniques. There are, however, several productions where this is not so. The result was that a degree of nondeterminism arose in the parser design. The authors attempted to solve this problem through a combination of left factoring and the employment of a simple single token look–ahead. This solution worked for all but the two productions described below. In one case a two token look–ahead was employed and backtracking was used in the other. This is not to say that the authors are absolutely certain that PHI is not an LL(1) grammar or that backtracking had to be used. These solutions were used because they solved the problem at hand.

A two token look–ahead was used for the production\(^\text{13}\) ARGBINDING = [ QUALEXP OP ]. When the token '[' is found, a flag is set to indicate that an ARGBINDING is being parsed. The first look–ahead token is utilized when parsing the QUALEXP part. QUALEXP.

\(^{13}\)A complete description of the PHI grammar may be found in the Appendices
for example, may parse as TERM, which in turn may parse as either FACTOR or FACTOR*TERM. After succeeding on FACTOR, a look-ahead is employed to look for the MULOP, *, to see if a recursive search for another TERM should be initiated. This methodology works as long as QUALEXP was not called from ARGBINDING. If it was called from ARGBINDING, argbinding flag set, the operator * could be the trailing operator in the ARGBINDING production and not part of the TERM production. In order to make this determination, an additional look-ahead is utilized to look for the token '}'. If '}' is found the QUALEXP production is terminated, e.g., term does not recursively call itself again, and the ARGBINDING production is allowed to proceed to completion.

Backtracking was utilized when parsing productions of ACTUAL: ACTUAL = COMPOUND and ACTUAL = DENOTATION = FORMALS I-> ACTUAL. Legitimate PHI sentential forms produced by the production FORMALS = (FORMALS+) are proper subsets of the sentential forms produced by the production COMPOUND = (ELEMENTS), excluding the empty compound statement. Since any number of identifiers may appear between the parentheses, it is not practical during the parse to utilize look-ahead to determine the presence of the token "|->". In effect, the parser first realizes it was parsing a DENOTATION when it finds "|->". This problem was solved by designing the parser to apply first the compound production when presented with this choice. If "|->" is later found, the parser then backtracks\(^{14}\) to the FORMALS production. The normal costs associated with backtracking were not evident in this isolated case. As described below, space trade-offs had previously been made and the parser was already working with an abstract syntax tree. The root to the subtree containing the previously parsed compound was simply passed to the FORMALS production to insure that the string could have been

\(^{14}\) A purist would say that this instance of backtracking means that the PHI compiler does not in fact employ a recursive-descent parser.
produced by FORMALS. After ascertaining FORMALS, the parser now continues the parse using the DENOTATION production.

The production QUALEXP = QUALEXP WHERE AUXDEFS required a deviation from pure recursive descent parsing. The semantics of this production are such that a terminal (e.g., an identifier) may be used prior to its definition. In itself, this does not present a major problem for the compiler writer. However, this construct also changes the scope of the identifier since the inner-most scope, in the form of the QUALEXP, is parsed first and the parser then works its way to the outer-most scopes, the AUXDEFS. This problem is analogous to that of mutual recursion in Pascal, without the benefit of the forward declaration [Ref. 4:p. 213].

Originally, the parser was designed to output the parse tree in flattened form, essentially a post-order walk of the tree. This design implemented traditional symbol-table management routines. However, after obtaining a clearer understanding of the semantics involved with the problems mentioned earlier, notably the production QUALEXP = QUALEXP WHERE AUXDEFS, the authors realized a traditional symbol-table would be too inefficient. Management of the table would take an inordinate amount of assets and be too unwieldy to work with. The authors solved this problem by maintaining the status of the parse in an abstract syntax tree so the output from the parser is now in tree form. This permits information originally held in the symbol-table to be maintained in the tree itself. The parser is able to analyze the source program by walking the tree and decorating the nodes with required information. Maintaining a binary tree in memory does require more space, but this is insignificant when compared with the benefits.

Interestingly, maintaining the parse in tree form presented several additional benefits. The solution to the aforementioned problem of distinguishing between COMPOUND and DENOTATION became trivial because it was now simply a matter of returning to the appropriate subroot and rewalking the tree. Also, working with a binary tree permitted the
authors to perform a modicum of optimization in the parser. It becomes relatively straightforward to perform compaction on an actual tree.

The authors think that this design offers maximum potential for future enhancements of the PHI programming environment. One possibility would be to use this front-end to drive a PHI interpreter. Modularization of the front-end in this manner simplifies functional understanding of the front-end and should lead to increased ease of maintenance and portability. To demonstrate portability, the authors recompiled the front-end and executed it on a 68000 based processor. This was accomplished with no modifications to the source program, just replacement of C run-time header files for the new target machine.

C. ERROR HANDLING

Tremblay and Sorenson [Ref. 3:p. 183] classify error responses into three categories:

I. Unacceptable responses
   1. Incorrect responses (error not reported)
      a. Compiler crashes
      b. Compiler loops indefinitely
      c. Compiler continues, producing incorrect object program
   2. Correct (but nearly useless)
      a. Compiler reports first error and then halts

II. Acceptable responses
   1. Possible responses
      a. Compiler reports error and recovers, continuing to find later errors if they exist
      b. Compiler reports the error and repairs it, continuing the translation and producing a valid object program
   2. Impossible with current techniques
      a. Compiler corrects error and produces an object program which is the translation of what the programmer intended to write

In the prototype PHI compiler, the authors have implemented a limited form of error recovery. The primary benefit of error recovery is to "prolong the compilation life of the program as long as possible before the compiler gives up on the source program". [Ref. 2:p. 11] This allows the maximum number errors to be discovered per compilation, shortening the edit, compile, debug cycle inherent to writing computer programs.

The authors analyzed the intended environment and use of the PHI compiler and decided that lexical analysis and syntactic analysis were the most likely source of errors.
Lexical errors basically involve invalid characters or incorrect tokens. Common examples of these types of errors are unrecognized words, misspelled identifiers/keywords, or illegal characters. Syntactic errors relate to incorrect structure of the program. These errors arise when the programmer failed to follow the rules, productions, of the grammar. The form of the program is wrong. [Ref. 9:p. 226, Ref. 3:p. 185]

One thing the error handler should not do is exacerbate the situation by reporting bogus errors or executing an erroneous program. To insure erroneous programs are not executed, the authors inhibited object file production if any errors were discovered. The authors do not believe the compiler should allow code generation to continue, or even begin, if the source program has errors. Often times one error leads to an avalanche of errors being reported and this is extremely annoying to the programmer. The authors attempted to minimize this situation, but found it impossible to eliminate completely because some errors feed on others. To insure the programmer would not become overwhelmed with error messages, the authors terminate the compilation after 10 errors. Also, for programmer convenience, actual error messages are outputted instead of error codes. The authors saw no justification in using a cryptic code when a plain language message served much better. Since the authors anticipate students in functional programming classes to be primary users of the PHI compiler, error messages have their basis in the productions describing the PHI language. It is assumed that users of the PHI compiler have an understanding of PHI's syntax.
III. BACK-END OF THE COMPILER

A. OVERVIEW

The back-end of the compiler consists of the semantic checker and code generator. Semantic checking and code generation are completed in one pass, and the output is a sequence of bytes, held in memory, which correspond to ASCII characters. These characters are then written to a text file, which the assembler uses to output an object file. This output is linked to the appropriate run-time routines to make a usable program. For the current implementation, a RASM86 assembler and LINK86\textsuperscript{15} linker are used.

B. RUN-TIME ORGANIZATION

Since PHI is a structured language with scoping and function calls, it lends itself to a stack-oriented run-time architecture. The stack is set up to accomplish two tasks: 1) to hold pointers to the current operands, and 2) to hold activation records for functions currently in use. Both of these tasks are described below.

There is a 64 kilobytes limit on memory used while a program is running. This limitation is imposed because the memory is addressed as an offset from a base address, and the maximum offset is 64K. This space is competed for by the stack, current variables, and constants (see Figure 3.1). The stack grows from the top of this space down, and the variable space grows from the base of this space up, preventing wastage by either component. Because PHI is a functional language, a value is returned from each operation, and a pointer to this value is placed at the top of the stack. The returned value is placed in the lowest available space in the part of memory assigned to variables and constants. A heap allocation method is not currently used because 1) all data types currently implemented use only one word of memory, and 2) there is no fragmentation of

\textsuperscript{15}RASM86 and LINK86 are trademarks of Digital Research, Inc.
memory because all types are currently static. If the next operation is a binary operation, a pointer to the second operand is placed on the stack, and the operation takes place using the two topmost pointers. The result is placed in memory, and the process begins afresh with new operands. If the next operation is unary (such as the negation operation), no change to the stack takes place and the variable in memory is altered as the program directs.

![Memory Organization](image)

**Figure 3.1**
Memory Organization

If the second operand of an operation is to be the result of a function call (e.g., "2 * f(x)"), an activation record is placed on top of the pointer to the first operand and the function's value is calculated. Then, the activation record is deleted and a pointer to the function result is saved and placed at the top of the stack.
The activation record itself, Figure 3.2, contains three parts: the static link, the static nesting level, and a pointer to the address in memory where the function's first variable is stored. The static link is a one-word pointer which points to the static nesting level space of the previous activation record, and is used to traverse the stack from activation record to activation record, i.e. a static chain [Ref. 4: p. 77]. The static nesting level and the pointer to the base of the storage space for a scope's values are used to access variables and constants. In this design, a two-tuple \((B, L)\) is associated with each variable. In this two-tuple, \(B\) represents the static nesting level and \(L\) is the offset within that level. By following the static chain for \((\text{current nesting level} - \text{target nesting level})\) links, the activation record of the scope of the target value can be accessed. Then, the address of the variable is calculated by adding \(L\) to the low address of the scope's variables. An alternate method would have been to store the values directly in the stack between or within activation records. However, this is a messy process when dealing with dynamic data structures such as sequences. Additionally, it is conceptually easier to divide the stack and the variables.

Functions are implemented as calls to assembly language subroutines, with pointers to the arguments placed on the stack before calling the routine. Using this scheme and noting the fact that PHI cannot have side effects, the implementation of recursion is straightforward. Whenever a function is called, its activation record is placed on the stack and pointers to its arguments are placed on top of the activation record. If the function is
recursive, the assembly language subroutine simply calls itself until the base of its recursion is reached or until stack overflow is reached. Figure 3.3 shows an example of a series of activation records called by a program with a recursive function. Note that the data definition ("answer") has no arguments and simply calls the factorial function. The factorial function, on the other hand, has an argument and it uses that argument as an operand. So, a pointer to that value is put on the stack and the next operand, fac (n - 1), is put on the stack as an activation record. When fac (n - 1) is evaluated, a pointer to its return value is placed on the stack. This cycle of evaluation, pop activation record, evaluation will continue until the data definition "answer" is evaluated.

As an example of the code generated for function calls and recursion, the following PHI program fragment is used:

\[
C(n) = \begin{cases} 1 & \text{if } n = 0 \\ n \cdot C(n-1) & \text{else} \end{cases}
\]

Figure 3.3
Factorial Program and Activation Records

28
This, of course, simply calculates the factorial of the integer \( n \). Figure 3.4 is the listing of the assembly language segment which is generated from this fragment.

<table>
<thead>
<tr>
<th>Address/Machine Code</th>
<th>Assembly Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>0103 E94A00</td>
<td>0150 jmp a10000</td>
</tr>
<tr>
<td></td>
<td>a10001:</td>
</tr>
<tr>
<td></td>
<td>mov cx,0</td>
</tr>
<tr>
<td>0106 B90000</td>
<td>E call i_formal</td>
</tr>
<tr>
<td>0109 E80000</td>
<td>mov ax,0</td>
</tr>
<tr>
<td>010C B80000</td>
<td>call iputvalue</td>
</tr>
<tr>
<td>010F E80000</td>
<td>call iequ</td>
</tr>
<tr>
<td>0112 E80000</td>
<td>call igetvalue</td>
</tr>
<tr>
<td>0115 E80000</td>
<td>cmp ax,1</td>
</tr>
<tr>
<td>0118 3D0100</td>
<td>0126 jne a10003</td>
</tr>
<tr>
<td>011B 7509</td>
<td>mov ax,1</td>
</tr>
<tr>
<td>011D B80100</td>
<td>call iputvalue</td>
</tr>
<tr>
<td>0120 E80000</td>
<td>014C jmp a10002</td>
</tr>
<tr>
<td>0123 E92600</td>
<td>a10003:</td>
</tr>
<tr>
<td></td>
<td>mov cx,0</td>
</tr>
<tr>
<td>0126 B90000</td>
<td>E call i_formal</td>
</tr>
<tr>
<td>0129 E80000</td>
<td>mov cx,0</td>
</tr>
<tr>
<td>012C B90000</td>
<td>E call i_formal</td>
</tr>
<tr>
<td>012F E80000</td>
<td>mov ax,1</td>
</tr>
<tr>
<td>0132 B80100</td>
<td>E call iputvalue</td>
</tr>
<tr>
<td>0135 E80000</td>
<td>E call isub</td>
</tr>
<tr>
<td>0138 E80000</td>
<td>E call ppop</td>
</tr>
<tr>
<td>013E 51</td>
<td>push cx</td>
</tr>
<tr>
<td>013F 57</td>
<td>push dx</td>
</tr>
<tr>
<td>0140 BB0100</td>
<td>mov bx, 1</td>
</tr>
<tr>
<td>0143 E80000</td>
<td>E call i_mov</td>
</tr>
<tr>
<td>0146 E8BDFF</td>
<td>0106 call a10001</td>
</tr>
<tr>
<td>0149 E80000</td>
<td>E call imult</td>
</tr>
<tr>
<td></td>
<td>a10002:</td>
</tr>
<tr>
<td>014C E80000</td>
<td>E call del_scope</td>
</tr>
<tr>
<td>014F C3</td>
<td>ret a10000:</td>
</tr>
</tbody>
</table>

Figure 3.4
Assembly Language Output from Factorial Program

The label "a10001" at address 0103 is the label of the subroutine which returns the factorial. When it is called, pointers to the values of the arguments are placed on the stack. If the subroutine is called before the base of the recursion is reached, a jump is made to label a10003. Then, the new actual value \((n - 1)\) is calculated and placed in the low part of memory, a pointer to the value is put on the stack, and the values are prepared for calling...
by the next subroutine (lines 0126 to 0143). The factorial subroutine is then called again. This process continues until the base of the recursion is reached; in this case a pointer to the integer value is put at the top of the stack (line 011D), and a jump is made to label a10002. Here, the subroutine "del_scope" tears down the activation record on the stack and puts a pointer to the result of the function at the top of the stack. Clearly, recursion in the PHI program can be implemented by a parallel recursion in the assembly language output of the compiler.

Another feature of the output code shown in Figure 3.4 is that there is an unconditional jump around the function (lines 0103 and 014F). This is a result of the decision to output inline code in spite of the fact that functions can be called at random. There are both space and time penalties to be paid for these jumps, especially since each function must have a jump and label instruction bracketing it. However, the ultimate effect of all these jumps is to get to the label at the bottom of the program. The result is that all but one jump/label pair could be eliminated by an optimizer, making the penalty trivial. Another solution considered was to generate code for functions and the "main" program separately, then combine the two when printing the output from the code generator. This was not done for reasons put forth in the section that describes the semantic analyzer.

Variable and constant storage is word oriented rather than byte oriented to take advantage of the 8086 processor's 16 bit capability. Integers and naturals are both represented as single words, and booleans are represented as integers, either 1 or 0. While this boolean representation is somewhat wasteful in terms of memory space, it allows for a great deal of overlapping in certain subroutines used in function calling and comparisons. It is planned to represent real numbers with two words of memory, and sequences using linked lists. Neither of these types have been fully implemented; however, there are provisions in the compiler for adding these features at a later date.
There is currently no dynamic allocation of registers. Some registers are used for specific purposes; for instance, the SI register is used to mark the top of the program stack, and of course the BP and SP registers are used to manage the machine's stack. In general, arithmetic processes take place in the AX register, using other general registers as auxiliaries as needed. When variable space is needed, the highest unused address space is allocated and, when a function is finished, only the result is saved in storage; all other value spaces are returned for use by the program.

Error handling is probably the simplest part of the run-time routines. Any run time error such as overflow or division by zero errors will result in an appropriate error message to the user (see Appendix O for a full listing of error messages). Then, program execution will terminate and control is returned to the operating system.

C. SEMANTIC CHECKING and CODE GENERATION

The PHI compiler utilizes the recursive descent technique to perform semantic checking and code generation in one traversal of the parser tree. In most cases, tree nodes are filtered through the semcheck function, which calls various procedures based on the name of the node. These procedures, in turn, call semcheck for each of their children, and the process is repeated until the leaves of the tree are reached. The function semcheck then returns a type (e.g., integer, real, boolean), which the parent node uses to determine the semantic correctness of its subtree. With the information returned from the semcheck function, the parent procedure can do one of three things: return a type, convert one node to a different type, or declare an error condition.

Concurrent with semantic checking, code is generated. As noted above, this is assembly language code written to a buffer in memory. If an error condition is declared, however, a flag is set and code generation ends. Semantic checking will then continue until the tree is completely traversed or ten errors are accumulated; then, the semantic checking
process terminates. Unlike the parser, the semantic checker makes no attempt at error recovery; top-down checking simply continues normally from where the error was detected.

Top-down semantic checking results in a neat, trim package for the back end of the compiler. Unfortunately, there are some problems that pure top-down checking will not solve. For instance, determining if there is a one-to-one match between formals and actuals for a given function involves some detours from top-down checking, as explained below.

The scoping rules of PHI provided the largest challenge to writing the semantic checker. One solution is a multiplicity of stacks. The size of these stacks depends upon the number of its constituents visible at any one time. Usually, the proper match for an item is the one found closest to the top of the stack. However, because of the semantics of the "and" construct, checks against the variable-stack do not always follow this convention.

There are four stacks used by the semantic checker: the type-stack, the variable-stack, the definition-stack, and the and-stack. All but the type-stack are implemented as linked lists. This implementation sheds the disadvantage of static length arrays at the cost of a slight increase in memory and temporal resources. The type-stack uses a fixed-length array of 300 entries because 1) the basic types of real, boolean, integer, natural, and trivial will be accessed most frequently, because they are the building blocks of every type and sequence, and because they can be more easily accessed from an array than from a linked list, 2) a list of 300 type entries should not impose an extreme burden on the programmer, and 3) the planned implementation of sequences will be more straightforward if the type-stack is an array.
The type-stack, Figure 3.5, is meant to hold both the basic type definitions and user defined type definitions. This stack holds both the name of the type and the number of bytes needed in memory to implement the type. At compiler initialization, it contains the five basic types and user defined types are added as they are encountered. The begin–end construct of the language (not implemented yet) allows declared types to be visible over a specified range. It is planned to implement this construct by setting a pointer to the top of the stack upon encountering the begin node and then popping the stack to that point after both of the node's subtrees have been checked.

The variable-stack, Figure 3.6, holds all of the variables, including function names, currently seen by the semantic checker. Each entry holds a pointer to the hash table containing labels, a type, a pointer to the tree node defining it, and a flag to designate whether or not it is a formal. Whenever a variable name is encountered and the name is not a call to a function and not a data definition, it is put into the variable stack. Then, when a scope is exited, the variables local to that scope are dropped from the stack. For example, after a function is defined, all of its formals are popped from the stack.
The definitions-stack, Figure 3.7, contains all of the function and variable definitions visible in a given scope; e.g., the declaration $C : R * Z \rightarrow B$ would put the definition $C$ into the definition-stack. This entry would contain the type of $C$'s return value (Boolean), a pointer to the tree node that contains $C$, and a pointer to a linked list which contains its argument types (Real and Integer). This last field will be null if the declaration is a data definition. This stack grows and shrinks in the same way as the type stack.

The authors considered combining the definitions-stack and the variable-stack because of the similarity between their fields. In fact, one of the primitive implementations was designed in this way. However, this slowed down the search for both definitions and variables considerably, and the overhead needed to implement these two as separate stacks is small: three extra functions and one extra pointer.

The need for the and-stack is derived from the scoping rules imposed by the AND construct. This construct allows a variable to be referenced before it is declared without the benefit of Pascal's forward declaration or equivalent. This is true of other constructs in PHI such as the WHERE construct. However, the AND construct cannot be parsed in such a way that the semantic checker can see all variables before they are used, because either subtree of the AND statement can define variables used by the other subtree. So, a program such as the one depicted in Figure 3.8 needs a vehicle by which it can detect that the variable $d$ is defined later in the program. The and-stack is such a vehicle.
Figure 3.8
Tree With Forward Variables

When the semantic checker reaches the AUXAND node, Figure 3.8, a flag is set to indicate that AUXAND has been traversed, and a pointer is set to the top entry of the and-stack. "Notfound" is returned from the semcheck function when the variable d is reached, but, since the AND condition has been set, a pointer to d is put in the and-stack. Note that d is later defined in a data definition (DATAUXDEF node), and when both the left and right subtrees of AUXAND have been checked, all variables in the and-stack are checked against variables in the variable-stack. If a match is found, d is defined and removed from the and-stack. In the event that a variable is not found when the AUXAND node's complete subtree has been checked, an error condition (UNDEFINED VARIABLE) would be set. The semantic checker would recognize this condition because the top of the and-stack would not be equal to the mark placed at the top of the stack when the AUXAND node was entered. Nested AUXANDS are possible, but they pose no problem because the top of the and-stack is marked when the auxand node is traversed.
Variables and functions are represented in the run-time by a call to an assembly language subroutine, and each subroutine must have a discrete name. Also, there are several labels found throughout the program, and each of these must have a name. These names are generated by the "name" function found in the \texttt{sem\_u\_c} module. Each name begins with the letter "a", followed by 6 digits. Examples can be seen in Figure 3.4.

Function definitions presented a problem that was solved with a deviation from pure top-down semantic checking. When a function definition (\texttt{FUNAUXDEF} in Figure 3.9) is encountered by semantic checker, the following procedure would be followed (see Figure 3.10 for the function definition entry):

\texttt{funid\_node:}

\begin{verbatim}
check for definition-stack entry for "f"
if not found
  return (ERROR)
get a pointer to the first formal of f
get a pointer to the first formal of definitions-stack entry
while both pointers <> Null do
\end{verbatim}

\begin{figure}
\centering
\includegraphics[width=0.5\textwidth]{tree.png}
\caption{Tree for Function $f$}
\end{figure}

\begin{equation}
 f (x,y) = x \times y
\end{equation}
put variable in varstack; use type pointed to by the formal list
advance both pointers
end while loop
if not (both pointers == nil)
  return (FORMALS MISMATCH)
else
  put "f" in the variable-stack
  return (Type of f = INTEGER)
end else
end.

funauxdef node:
  left type = semcheck (Left Child)
  right type = semcheck (Right Child)
  if (left type <> right type)
    call a procedure which will either
    convert the right type to the left type or set an error flag.
  endif
end.

When a function is called with arguments, a similar process takes place (refer to Figure 3.11):

actualist: Input is a pointer to the actualist node
  Output is error condition
  Check definitions-stack for "f"
  if "f" not found
    set error (FUNCTION DEFINITION NOT FOUND)
  set elistptr to first element of element list
  elist (elistptr)
  check var stack for "f"
  if found,
    generate code to call "f"
  if not found
    if and_flag = TRUE
      put "f" in the and stack
    else
      set error (FUNCTION NOT DEFINED)
end.

elist: Input is a pointer to the element list node
  if pointer->rptr <> nil
    elist (pointer->rptr)
  check type of element against corresponding formal type
  if types don't match
    set error (IMPROPER ARGUMENT TYPE)
else

generate code to put pointers to argument values on the run-time stack
end.

Figure 3.10
Definitions—Table Entry For Function f

Type conversions are implemented in the semantic checker, albeit the code generator
does not yet support this feature. The function hnumconvert (half number-convert,
found in the module sem0) will check to see if a conversion of the right subtree of a node
to the left subtree type should be accomplished. This is useful for function definitions,
where the body of the function may be converted to the type the function returns, but the
converse is not acceptable. In addition, the function numconvert (found in the sem0
module) will convert either the left tree type or the right tree type of a node. This is useful
for certain arithmetic operations. The semantic checker considers integer-to-real and
natural-to-real conversions to be legal. Natural to integer conversions are not implicitly
done, since both of these types are represented in exactly the same way. On the other
hand, an attempt to return an integer value for a function which has a declared type of
natural will result in an error.
Variables of simple type (i.e., natural, integer, or real) need not be declared before use, although such a declaration may be made. If a variable is undeclared when defined by a data definition, the semantic checker will attempt to classify it. If the semantic checker expects to find a boolean value, the variable is easily classified as a boolean and an entry is put into the variable table. If a numeric variable is expected, the semantic checker will try to type it as an integer; failing this, it will be classified as a real number. However, the AND construct alters this somewhat. If a variable is used before it is defined by a data definition, it must have been defined using the LETDEF construct.

As noted in the section on run-time, some thought was given to generating all functions and data definitions to one buffer and the "main" program which calls these functions to another buffer. However, this would be an inefficient use of memory space.
since one buffer might run out of space while the other is under-utilized. Although there is
a proliferation of jump calls in the output using one buffer, an optimizer could easily
eliminate all but one call, as noted above.

D. OPTIMIZATION

There is no optimization module implemented in the PHI compiler. In this section an
attempt will be made to identify three types of optimization which are suitable for
implementation. Also, a small dissertation on what optimizations should not be considered
is included.

The first suitable type of optimization is constant folding. The purpose of constant
folding is to eliminate multiple consecutive constants in arithmetic expressions [Ref 3:p.
612], and the function numconv in module sem0 makes an excellent structure in
which to implement this optimization. This is because most arithmetic operations call this
function. It would be straightforward to put a function that tests the left and right children
of an operand node to see if they are constants, then perform the operation in the compiler
and generate code for a constant call. However, since the division operators do not call
numconv, the constant folding function would have to be inserted in idiv and rdiv
also.

The other two optimizations are post-code generation optimizations. The first one
considered is jump optimization. This should be the most worthwhile to implement: if the
number of functions and data definitions is n, n > 0, there will be n - 1 unnecessary
unconditional jump statements and labels.

These jump statements can be eliminated by replacing the first "jmp" statement with a
jump to the last label in the code; then, because "jmp" is not used for anything except to
circumnavigate functions and data definitions, all other unconditional jumps and their labels
can be eliminated.
The last type of optimization is a form of peephole optimization. Occasionally, there will be a 'call push' statement followed by a 'call pop' statement. This is unnecessary and can be eliminated. The 8086 assembly code equivalent of push, followed by pop should not occur in the present design.

Dead code optimization eliminates code inside a jump when that code contains no labels. It is not necessary to implement this type of optimization with the current design, since unconditional jumps are only used to bracket functions and definitions. However, if one accepts the premise that programmers occasionally make mistakes, it might be worthwhile to keep track of which functions are called and eliminate code for those which are not. A message to the programmer concerning this circumstance would be useful, too.
IV. RESULTS & CONCLUSIONS

A. RESULTS

The implementation described in this study demonstrates the design and implementation of a compiler for the functional programming language PHI. Since this implementation is a prototype, it does not possess all of the qualities desirable in a full implementation. However, the necessary hooks are present and the design is mature enough to allow expanding the prototype to a full implementation.

The PHI compiler front-end implements machine independent lexical and syntactic analyzers. This implementation is complete and faithfully follows the syntax of PHI — based on the design of the language as of 07 January 1987. In deciding which modules to include in the front-end and back-end, the authors were originally guided by the traditional methodology of placing the analysis functions in the front-end and generative functions in the back-end (Ref. 8:p. 201). However, as the design of the PHI compiler progressed, the authors removed semantic analysis from the front-end and combined it with code generation. This produced a one-pass semantic analysis/code generation phase.

The PHI compiler back-end implements a machine dependent one-pass semantic analyzer and Intel 8086 code generator. The semantic analyzer implements the basic constructs of PHI. Functions and data definitions may be defined, and the integer, natural number, real number, and boolean types are fully implemented. Implementation of code generation is congruent to that of the semantic analyzer, with the exception that the real number data type has not been implemented.
B. CONCLUSIONS

It is possible, using traditional technologies to design and implement a compiler for the functional programming language PHI. It is not possible to utilize either pure recursive descent or pure deterministic techniques for this implementation. The syntax/semantics of the language forced a degree of non-determinism, and one instance of back-tracking was required in the PHI compiler front-end.

The overall design is highly modularized facilitating the understanding of concept and implementation. The authors think that this approach will greatly reduce maintenance costs and provide greater flexibility in making changes and additions to the PHI programming environment. It should be possible, for example, to use the front-end described in this thesis to drive a PHI interpreter. Being able to abstract out this front-end and use it without change should make the implementation of a PHI interpreter relatively simple. Modularizing the design also increases portability of the compiler to other machines. To demonstrate portability, the authors recompiled the front-end and executed it on a 68000 based processor. This was accomplished with no modifications to the source program, just replacement of C run-time header files for the new target machine.

Removing the semantic analyzer from the front-end permitted coupling semantic analysis with code generation. The fixed-length buffer design of the code generator is suitable for this prototype implementation but should be redesigned utilizing dynamic buffer allocation methods in follow on implementations. The authors think that utilizing a single pass through the parse tree is practical for the basic constructs of PHI and believe this methodology is suitable for future designs of the PHI compiler.
V. FURTHER RESEARCH

Further research may be broken down into two major areas: short and long range projects. The former may be further broken down into two main areas: adding unimplemented features and improving the PHI programming environment. On the other hand, all long-range projects involve only the programming environment. All of these areas are discussed below.

In the prototype of the PHI compiler, both Real and Compound variable types remain unimplemented. Compound variable types consist of sequences, the Trivial type, user defined types, and tuples. Although all of these are recognized by the parser, the semantic checker will not recognize complex types and no code will be generated. The Real type is recognized by the semantic checker, which can discern if conversion from an integer or natural type should be accomplished; however, no code is generated to implement this type in the run-time structures. Note also that operators which operate solely on complex types and reals (e.g., the real divide and concatenate operators) are not implemented.

One other operator not implemented is the "|->" operator. In addition, argument bindings, functionals, and FILEs are not recognized by either the semantic checker or the code generator.

Short-range improvements to the PHI environment may come either after a full implementation is accomplished or may be developed concurrently with the full implementation. Admittedly, the current environment is analogous to instrumentation on a helicopter: there is just enough to know that the system is running! The environment could be improved by implementing the interactive mode of PHI, as opposed to the current batch mode. A sample interactive session of PHI may be found in [Ref. 1:pp 1-17]. Also, an interpreter would be a good starting point toward developing a practical, working
environment for PHI. As noted above, the front end of the prototype compiler may be adapted for this purpose; alternatively, due to the structural similarities between PHI and LISP, an ambitious researcher may wish to write an interpreter in LISP.

One final short-range improvement which is not covered by either category would be to allow more than 64K of run-time memory. It would be worthwhile to take advantage of the large amount of memory most modern microcomputers have, especially since sequences and recursion, upon which PHI is based, gobbles up memory with abandon.

When the PHI compiler becomes a serious user's tool, some long-range research will become viable. Sophisticated input and output would be a vital consideration, and the minimal I/O methods now in use would need substantial improvement. The most ambitious researchers in this direction should consider a bit-mapped display with the possibility of a syntax-directed editor. Also, based on the authors' limited experience in PHI programming, a debugger would be a necessary tool for the serious programmer.
LIST OF REFERENCES


APPENDIX A

THE FUNCTIONAL LANGUAGE PHI — Φ

(CONCRETE SYNTAX OF Φ — 10/16/86)

Grammatical Notation:

Both '{C₁, C₂, ..., Cₙ}' and \( \{ C_1 \} \) mean exactly one of C₁, C₂, ..., Cₙ.

Similarly, '[C₁ | ... | Cₙ]' and \[ C_1 \] mean at most one of C₁, ..., Cₙ. The notation 'C^*' means zero or more Cs; 'C^+' means one or more Cs; 'CD ...' means a list of one or more Cs separated by Ds. Terminal symbols are quoted when they could be confused with metasymbols.

Grammar:

\[
\begin{align*}
\text{BLOCKBODY} & = \{ \text{QUALEXP} \} \hspace{1em} \text{LET DEFS} ; \text{BLOCKBODY} \\
\text{DEF} & = \{ [\text{ID}] \text{FORMALS} = \text{QUALEXP} \} \\
& \hspace{1em} \text{ID} : \text{TYPEEXP} \\
& \hspace{2em} \text{TYPE ID [FORMALS]} = \text{TYPEEXP} \\
\text{QUALEXP} & = \{ \text{EXPRESSION} \} \hspace{1em} \text{WHERE AUXDEFS} \\
\text{AUXDEFS} & = \text{AUXDEF AND} ... \\
\text{AUXDEF} & = [\text{ID}] \text{FORMALS} = \text{EXPRESSION} \\
\text{FORMALS} & = \{ \text{ID} \} \\
& \hspace{1em} \{ ( \text{FORMALS}, ... ) \} \\
\text{EXPRESSION} & = \{ \text{EXPRESSION} \} \text{\lor CONJUNCTION} \\
\text{CONJUNCTION} & = \{ \text{CONJUNCTION} \} \text{\land NEGATION}
\end{align*}
\]
NEGATION = \{ \sim \} \text{RELATION}
RELATION = \{ \text{SIMPLEXP RELATOR} \} \text{SIMPLEXP}
RELATOR = \{ \leq, \geq, <, > \}
SIMPLEXP = \{ \text{SIMPLEXEP ADDOP} \} \text{TERM}
ADDOP = \{ +, * \}
TERM = \{ \text{TERM MULOP} \} \text{FACTOR}
MULOP = \{ / \}
FACTOR = \{ \text{primary} \}
PRIMARY = \{ \text{APPLICATION PRIMARY APPLICATION} \}
APPLICATION = \{ \text{APPLICATION ACTUAL} \}
ACTUAL = \{ \text{ID DENOTATION COMPOUND ARGBINDING BLOCK FILE CHAR} \}
DENOTATION = \{ \text{CHAR* DIGIT+, DIGIT* FORMALS} |\to \text{ACTUAL} \}
CONDITIONAL = \text{IF ARM ELSIF ... [ELSE EXPRESSION] ENDIF}
ARM = \text{EXPRESSION THEN EXPRESSION}
COMPOUND = \{ \text{ELEMENTS} \}
ELEMENTS = \{ \text{QUALEXP, ...} \}
ARGBINDING = \{ [\text{OP QUALEXP}] \text{OP QUALEXP} \}
OP = \{ +, \times \}
BLOCK = \text{BEGIN BLOCKBODY END}
DEFS = DEF AND ...

TYPEEXP = TYPEDOM [ → TYPEEXP ]

TYPEDOM = TYPETERM [ + TYPEDOM ]

TYPETERM = TYPEFAC [ X TYPETERM ]

TYPEFAC = { TYPEPRIMARY
            TYPEPRIMARY∗
            ID << TYPEEXP, ... >>

TYPEPRIMARY = { ID
               PRIMTYPE
               ( TYPEEXP )

PRIMTYPE = { R | Z | N | B | I | TYPE }

For batch use, a program is considered a BLOCKBODY; for interactive use it is considered a SESSION:

SESSION = COMMAND∗

COMMAND = { DEF
            QUALEXP

            ;

49
APPENDIX B

THE FUNCTIONAL LANGUAGE PHI — Φ

(CONCRETE SYNTAX OF Φ — 03/03/87)

**Grammatical Notation:**

Both \( \{C_1, C_2, \ldots, C_n\} \) and \( \begin{cases} C_1 \\ C_2 \\ \vdots \\ C_n \end{cases} \) mean exactly one of \( C_1, C_2, \ldots, C_n \).

Similarly, \( \{C_1 \mid \ldots \mid C_n\} \) and \( \begin{bmatrix} C_1 \\ \vdots \\ C_n \end{bmatrix} \) mean at most one of \( C_1, \ldots, C_n \). The notation \( C^* \) means zero or more \( C \); \( C^+ \) means one or more \( C \); \( CD \ldots \) means a list of one or more \( C \)s separated by \( D \)s. Terminal symbols are quoted when they could be confused with metasymbols.

**Grammar:**

\[
\text{BLOCKBODY} = \begin{cases} \text{QUALEXP} \\ \text{LET DEFS} ; \text{BLOCKBODY} \end{cases}
\]

\[
\text{DEF} = \begin{cases} \text{REC} \begin{bmatrix} \text{[ID, ... ; TYPEEXP BE IS ] [ID] FORMALS = QUALEXP} \\ \text{TYPE ID [FORMALS] = TYPEEXP} \end{bmatrix} \end{cases}
\]

\[
\text{QUALEXP} = \begin{cases} \text{EXPRESSION} \\ \text{QUALEXP WHERE AUXDEFS} \end{cases}
\]

\[
\text{AUXDEFS} = \text{AUXDEF AND ...}
\]

\[
\text{AUXDEF} = \begin{bmatrix} \text{[ID] FORMALS = EXPRESSION} \end{bmatrix}
\]

\[
\text{FORMALS} = \begin{cases} \text{ID} \\ \text{(FORMALS, ...)} \end{cases}
\]

\[
\text{EXPRESSION} = \begin{bmatrix} \text{[EXPRESSION V] CONJUNCTION} \end{bmatrix}
\]

\[
\text{CONJUNCTION} = \begin{bmatrix} \text{[CONJUNCTION A] NEGATION} \end{bmatrix}
\]

50
NEGATION = [ ~ ] RELATION
RELATION = [SIMPLEXP RELATOR] SIMPLEXP
RELATOR = ( = | ≠ | > | < | ≤ | ≥ | ∈ | ∉ | → )
SIMPLEXP = [SIMPLEXP ADDOP] TERM
ADDOP = ( + | - | : | ^ | + | ' )
TERM = [TERM MULOP] FACTOR
MULOP = ( X | / | + | o | ; | X )
FACTOR = [ + | - ] PRIMARY
PRIMARY = { APPLICATION
{ PRIMARY APPLICATION }
APPLICATION = [APPLICATION] ACTUAL
{ ID [« TYPEEXP, ...» ]
DENOTATION
CONDITIONAL
COMPOUND
ARGBINDING
BLOCK
{ FILE | STREAM | ' CHAR' + }
DENOTATION = { ' CHAR' *
DIGIT + [ . , DIGIT + ]
NIL
FORMALS | → ACTUAL }
CONDITIONAL = IF ARM ELSIF ... [ELSE EXPRESSION] ENDIF
ARM = EXPRESSION THEN EXPRESSION
COMPOUND = { ‘ ELEMENTS ’}
( ELEMENTS )
{ ‘ ELEMENTS ’}
< ELEMENTS >
ELEMENTS = [ EXPRESSION, ... ]
ARGBINDING = [ { OP
OP ACTUAL }
ACTUAL OP ]
OP = [ , | RELATOR | ADDOP | MULOP | SUB ]
BLOCK = BEGIN BLOCKBODY END
DEFS = DEF AND ...
TYPEEXP = TYPEDOM [ \rightarrow TYPEEXP ]
TYPEDOM = TYPETERM [ + TYPEDOM ]
TYPETERM = TYPEFAC [ X TYPETERM ]
TYPEFAC = \{ TYPEPRIMARY
          TYPEPRIMARY [ ACTUAL ] \}
TYPEPRIMARY = \{ ID \{ \{ TYPEEXP, ... \} \}
PRIMTYPE = \{ ID \{ \{ TYPEEXP \} \}
PRIMTYPE = \{ R | Z | N | B | I | TYPE \}

For batch use, a program is considered a BLOCKBODY; for interactive use it is considered a SESSION:
SESSION = COMMAND+
COMMAND = \{ LET DEF QUALEXP \};
APPENDIX C
ASCII REPRESENTATION OF — Φ

<table>
<thead>
<tr>
<th>Reference</th>
<th>ASCII</th>
</tr>
</thead>
<tbody>
<tr>
<td>=</td>
<td>==</td>
</tr>
<tr>
<td>&lt;</td>
<td>LESS</td>
</tr>
<tr>
<td>≤</td>
<td>&lt;=</td>
</tr>
<tr>
<td>&gt;</td>
<td>&gt; &gt;</td>
</tr>
<tr>
<td>≠</td>
<td>IN</td>
</tr>
<tr>
<td>∈</td>
<td>NOTN</td>
</tr>
<tr>
<td>∨</td>
<td>∨</td>
</tr>
<tr>
<td>∧</td>
<td>∧</td>
</tr>
<tr>
<td>~</td>
<td>~</td>
</tr>
<tr>
<td>X</td>
<td>*</td>
</tr>
<tr>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>+</td>
<td>%</td>
</tr>
<tr>
<td>→</td>
<td>-&gt;</td>
</tr>
<tr>
<td>^</td>
<td>^</td>
</tr>
<tr>
<td>i&gt;</td>
<td>i&gt;</td>
</tr>
<tr>
<td>A_i</td>
<td>A ! i</td>
</tr>
<tr>
<td>T²</td>
<td>T @</td>
</tr>
<tr>
<td>R</td>
<td>$R</td>
</tr>
<tr>
<td>Z</td>
<td>$Z</td>
</tr>
<tr>
<td>N</td>
<td>$N</td>
</tr>
<tr>
<td>B</td>
<td>$B</td>
</tr>
<tr>
<td>l</td>
<td>$l</td>
</tr>
</tbody>
</table>
APPENDIX D
THE FUNCTIONAL LANGUAGE—Φ
(RIGHT-RECURSIVE GRAMMAR)

Note: 
(\ldots)^* \text{ means zero or more occurrences}
(\ldots)^+ \text{ means one or more occurrences}
(\ldots)^n \text{ means from zero to } n \text{ occurrences}
(x \mid y) \text{ means either } x \text{ or } y, \text{ but not both}

BLOCK ::= BEGIN BLOCKBODY END
BLOCKBODY ::= LET DEF; BLOCKBODY
DEF ::= DEF (AND DEF)^*
DEF ::= (ID)^1 FORMALS = QUALEXP
ID: TYPEEXP
TYPE ID (FORMALS)^1 = TYPEEXP
QUALEXP ::= EXPRESSION (WHERE AUXDEFS)^*
AUXDEFS ::= AUXDEF (AND AUXDEF)^*
AUXDEF ::= (ID)^1 FORMALS = EXPRESSION
FORMALS ::= (FORMALS (FORMALS)^1)
ID
EXPRESSION ::= CONJUNCTION (\lor CONJUNCTION)^*
CONJUNCTION ::= NEGATION (\land NEGATION)^*
NEGATION ::= (\sim)^1 RELATION
RELATION ::= SIMPLEXEP (RELATOR SIMPLEX)^1

54
RELATOR ::= = ≠ < ≤ ≥ ∈ ε

SIMPLEXP ::= TERM (ADDOP TERM)*

ADDOP ::= + - ^

TERM ::= FACTOR (MULOP FACTOR)*

MULOP ::= * /

FACTOR ::= + PRIMARY - PRIMARY PRIMARY

PRIMARY ::= APPLICATION (APPLICATION)*

APPLICATION ::= (ACTUAL)*

ACTUAL ::= ID

DENOTATION ::= (CHAR)^* CHAR can = ASCII 32-126

DIGIT ::= DIGIT^* DIGIT can = 0-9

FORMALS 1→ ACTUAL

ID ::= ALPHENUM^* ALPH can = a-z / A-Z / 0-9

CONDITIONAL ::= IF ARM | ELIF ARM | ELSE EXPRESSION | ENDIF

ARM ::= EXPRESSION THEN EXPRESSION
COMPOUND \( ::= \{(\text{ELEMENTS})^*\} \)
\( \{(\text{ELEMENTS})^*\} \)
\( <(\text{ELEMENTS})^* > \)

ELEMENTS \( ::= \text{QUALEXP},(\text{QUALEXP})^* \)

ARGBINDING \( ::= [ \text{op} ] \)
\( [ \text{op QUALEXP} ] \)
\( [ \text{QUALEXP OP} ] \)

OP \( ::= \) \( , \) \( \text{RELATOR} \)
\( \text{ADDOP} \)
\( \text{MULOP} \)

TYPEEXP \( ::= \text{TYPEDOM} ( \to \text{TYPEDOM})^* \)

TYPEDOM \( ::= \text{TYPETERM} (\to \text{TYPETERM})^* \)

TYPETERM \( ::= \text{TYPEFAC} (\to \text{TYPEFAC})^* \)

TYPEFAC \( ::= \text{TYPEPRIMARY} \)
\( \text{TYPEFAC} \)
\( \text{ID} \) \( <\text{TYPEEXP} (\to \text{TYPEEXP})^* > \)

TYPEPRIMARY \( ::= (\text{TYPEEXP}) \)
\( \text{ID} \) \( \text{PRIMTYPE} \)

PRIMTYPE \( ::= \mathbb{R} \)
\( \mathbb{Z} \)
\( \mathbb{H} \)
\( \mathbb{B} \)
\( \mathbb{I} \) \( \text{TYPE} \)

FOR INTERACTIVE IMPLEMENTATION OF \( \Phi \)

SESSION \( ::= (\text{COMMAND})^* \)

COMMAND \( ::= (\text{DEF QUALEXP}) ; \)
APPENDIX E
ROCK COMPILER HEADER FILES

/*
 * This file contains header files required by the Rock compiler
 */

/*
 * Name: scanner definitions
 * File: scanner.h
 * Authors: Maj E.J. Cole / Capt J.E. Connell
 * Started: 10/10/86
 * Archived: 12/11/86
 * Modified: 01/10/87 - Update keywords JC
 */

/*
 * This file contains definitions used by the scanner, parser, and
 * error recovery routines.
 */

/*
 * Modified: 01/10/87 Corrections to comply with latest definitions
 * of the language and update keywords. JC
 */

/*
 * General Token Types */
 * Listing of symbols can be found at end of list */

#define EOF 0
#define FALSE 1
#define TRUE 0
#define PYETERM 1
#define VAX KEYWORDS 17
#define NAMESIZE 18
#define VAXLINE 80
#define TABLESIZE 107

/* System dependent - sure! */
/* Real 14, 16 bit - don't! */
/* Length of string in chars - don't! */
/* Hash const size (same as...) */

57
define COMMA 17
define LTPAREN_ 18
#define RTPAREN_ 19
#define EQUIV_ 20
#define ORLOG_ 21
#define ANDLOG_ 22
#define NEQLOG_ 23
#define COLON_ 24
#define CAT_ 25
#define LTBRACKET_ 26
#define RTBRACKET_ 27
#define LTSQUIG_ 28
#define RTSQUIG_ 29
#define EMPT_LIT_ 30
#define RTARROW_ 31
#define LITERAL_ 32
#define IDENTIFIER 33
#define CONSTANT_ 34
#define REAL_ 35
#define INTEGER_ 36
#define NATURAL_ 37
#define BOOLEAN_ 38
#define TRIVIAL_ 39
#define CHAR_ 40
#define STRING_ 41
#define PCS_ 42
#define NEC_ 43
#define KW_ 44

#define EOF, error, unknown token, <=, >, <=, >=, +, -, *, /, :, !,
,, ( ), =, */", ", \, /\, ~, ":, , [ ], ], }, <->, literal,
identifier, constant, SR, SZ, SN, SB, S1, character, string, &,
unary plus, unary minus, keyword

#define AND 8
#define BEGIN 9
#define ELSE 10
#define ELSEIF 11
#define END 12
#define ENDF 13
#define FILE 14
#define GREATER 15
#define IF 16
#define IN 17
#define LESS 18
#define LET 19
#define NOTIN 20
#define READ 21
#define THEN 22
#define TYPE 23
#define WHERE 24
#define WRITE 25

/* Keywords */
#define CALLOC(y,x) ((y*) calloc(y,sizeof(x)))

struct NStruct {
    char NAME[NAME_SIZE];
    struct NStruct *line;
};
typedef struct NStruct NameRec;
extern char *callloc();
extern char *malloc();

#endif
PUBLIC DOMAIN SOFTWARE

Name: parser definitions
File: parser.h
Authors: Maj E.J. COLE / Capt J.E. CONNELL
Started: 10/20/86
 Archived: 12/11/86
 Modified: 01/12/87 - update NodeStruct definition JC

This file contains definitions used by the parser

Modified: 01/10/87 - update NodeStruct to hold the type of the node

```c
#ifndef LETDEF
#define LETDEF 1
#define DEFAND 2
#define KINDER 3
#define FUNID 14
#define UNDEF 15
#define DATADEF 16
#define TFDEF 17
#define TDFON 18
#define DATAANDDEF 19
#define FANDAGER 41
#define ANDAND 42
#define ACTUALIST 43
#define SEQUENCE 44
#define FORMAL 45
#define FLLIST 46
#define EMPTYCOMPONENT 48
#define EMPTYSEQUENCE 49
#define ANDAND 41
#define ANDAND 42
#define TYPESEQ 44
#define TYPEQUE 45
#define TYPEEXPLIST 47
#endif
```
typedef struct NodeStruct *NodeRec; /* global var - list number errors */
typedef struct NodeStruct *NodeRec, *nodal;

NodeRec *CreateNode();
node *NodeName();

extern int num_errors; /* during scan and parse */
extern int acquired; /* global flag - used to make Ptr */
extern int deterministic;

extern char *allocate();
extern char *malloc();
extern ErrorHandle();
extern WriteErrors();

/************************** External Utility Functions *******************/

extern NodeRec *CreateNode();
extern char *NodeName();
extern MakeNewRoot();
extern *form();
extern *BAll();
extern *Exist();
extern long ByPass();

#include <scanner.h>
#include <errors.h>

*end;
/*
*Name: error file definitions
*File: errors.h
*Authors: Maj E.J. COLE / Capt J.E. CONNELL
*Started: 01/20/87
*Archived: 04/07/87
*Modified:

This file contains definitions used by the error recovery routines.

Modified

 ifndef MAXERRORS
#define MAXERRORS 10

/******************** PARSE ERRORS ********************/

#define ERR0 0 /* ' ' or ' '-' w/o ' ' */
#define ERR1 1 /* ' ' or ' ' w/o '/' -- bad logical OR */
#define ERR2 2 /* '/' w/o '/ ' -- bad logical OR */
#define ERR3 3 /* ' $ ' w/o proper following char */
#define ERR4 4 /* invalid numeric constant */
#define ERR5 5 /* literal w/o ending */
#define ERR6 6 /* unidentified char in input file */
#define ERR7 7 /* out of memory */
#define ERR8 8 /* error in statement following */
#define ERR9 9 /* error in type definition following */
#define ERR_a 10 /* following 'xx' */
#define ERR_b 11 /* unable to complete eval of */
#define ERR_c 12 /* the blockbody */
#define ERR_d 13 /* missing or misplaced ; after */
#define ERR_e 14 /* definition */
#define ERR_f 15 /* invalid QualExp */
#define ERR_g 16 /* invalid TypeExp */
#define ERR_h 17 /* bad or missing forms */
#define ERR_i 18 /* missing or misplaced */
#define ERR_j 19 /* following 'xx' */
#define ERR_k 20 /* error in processing */
#define ERR_x 21 /* successive Actnals */
#define ERR_l 22 /* missing literal after keyword */
#define ERR_m 23 /* FILE */
#define ERR_n 24 /* missing or invalid exp after */
#define ERR_o 25 /* keyword */
#define ERR_p 26 /* IF statement w/o ENCLIF */
#define ERR_q 27 /* error in forms of precedence */
#define ERR_r 28 /* error in forms of precedence */
#define ERR_s 29 /* following comma or p */
#define ERR_t 30 /* error on ArgBinding - check */
#define ERR_u 31 /* QExp or */
#define ERR_v 32 /* off in DEQINE-implemuted */
#define ERR_w 33 /* feature */

#endif /* MAXERRORS */
/* NOTE: s through z reserved for future use */

/********************* SEMANTIC ERRORS **********************/

#define ERR_s 28 /* */
#define ERR_t 29 /* */
#define ERR_u 30 /* */
#define ERR_v 31 /* */
#define ERR_w 32 /* */
#define ERR_x 33 /* */
#define ERR_y 34 /* */
#define ERR_z 35 /* */
#define ERR_aa 35 /* Numeric value expected */
#define ERR_bb 35 /* Natural expected */
#define ERR_cc 35 /* Integer or natural expected */
#define ERR_dd 35 /* Error in tuple definition */
#define ERR_ee 35 /* Undefined var in "and" scope */
#define ERR_ff 35 /* Function w/o function def */
#define ERR_gg 35 /* Formal mismatch */
#define ERR_hh 35 /* Undefined function */
#define ERR_ii 35 /* Real number expected */
#define ERR_jj 35 /* Invalid constant */
#define ERR_kk 35 /* Boolean value expected */
#define ERR_ii 35 /* Boolean operator expected */
#define ERR_mm 35 /* Out of run-time memory space */

#endif
/**
 * PUBLIC DOMAIN SOFTWARE
 * 
 * Name : Semantic Definitions Header File
 * File : Semcheck.h
 * Authors : Maj E.J. COLE / Capt J.E. CONNELL
 * Started : 01/01/87
 * Archived : 04/10/87
 * Modified : 04/13/87 "FILENAME" eliminated EC
 *
 * This file contains the header file and definitions for the semantic
 * checker and code generator of the PHI compiler
 *
 * Modified : 04/13/87 "FILENAME" eliminated; output path now
 * depends on user's input EC
 */

/******* Externals ********/
#include <scanner.h>
#include <parser.h>
#include <errors.h>
#include <stdio.h>

/******* Globals ********/
#define NOTFOUND 0
#define UNTYPED 0
#define BOOLEAN 1
#define BOL_BYTES 2
#define REAL 2
#define REAL_BYTES 4
#define INTEGER 3
#define INT_BYTES 2
#define NATURAL 4
#define NAT_BYTES 2
#define ERROR 0
#define MAXADDR 64000
#define MAXTYPES 300
#define CODE_SIZE 20000
#define START_ADDR 0
#define TYPE_INIT 5
#define CNTRL 2 26
#define ENSTRING 0
#define NUM_BASE 48
#define STACKSIZE 10000
#define SIZBUFFER 30000
#define ADD 1
#define SUB 2
#define DIVIDE 3
#define MULT 4
#define SEM_ERR 0

#define NULL
#define NULL 0

endif
/* Type Definitions */

typedef int optype, FLAG, INTYPE;

typedef char sq [20];

typedef struct and struct *and_ptr;

/* Typetable Definitions */
typedef struct typenode {
    char name [10];
    int bytes;
    struct typenode *typeptr;
} tnode;

/* Formallist Definitions */
typedef struct formnode {
    int name, type;
    struct formnode *link;
} fnode;

/* Vartable Definitions */
typedef struct varnode {
    int type,
    form,
    def;
    nodal nptr;
    fnode *fptr;
    struct varnode *link;
} vnptr;

/* Deftable Definitions */
typedef struct defnode {
    int type;
    nodal nptr;
    fnode *fptr;
    struct defnode *link;
} dnptr;

/* And Definitions */

struct and struct {
    nodal ptr;
    int buffptr;
    struct and struct *link;
};

*/
#define BUFFLENGTH 30  /* Max size of input file name - directory */
#define NOTFOUND 0  /* Input buffer size */
#define BSIZE 1000  /* Input block size */
#define BLOCKSIZE 50  /* ASCII Equivalents */
#define BACKSPACE 8  /* ASCII Equivalents */
#define EOLN 13  /* ASCII Equivalents */
#define ESCAPE 27  /* ASCII Equivalents */
#define GETPROGRAM "Program to Compile -> "  /* Messages to observe */
#define HEADER1 "ROCK COMPILER"
#define HEADER2 "Press Escape Key to Exit Compiler"
#define FILE1_ERROR "File not Found"
#define FILE2_ERROR "Press ESCAPE to exit, any other key to continue"
#define WAIT "Compiling: Please Wait"
#define FAUSE "PRESS ANY KEY TO CONTINUE"
#define ERRORFILE "errors.phi"  /* Text file of errors */
APPENDIX F

ROCK COMPILER — MAIN MODULE

/* PUBLIC DOMAIN SOFTWARE */

/* Name : Main Rock Module */
/* File : Rock_main.c */
/* Authors : Maj E.J. COLE / Capt J.E. CONNELL */
/* Started : 01/06/87 */
/* Archived : 04/10/87 */
/* Modified : 04/13/87 Output files put to vdisk EC */

This file contains the following modules for the PHI compiler:

R_Initial  Semcheck  Main

Algorithm :

This contains the main procedure for the phi compiler, in addi-
tion to the initialization procedure & the main semantic checking
procedure. The main module inits the program, sets up the screen
by calling "user ()", & decides whether an error routine needs
to be called. It also closes out the input file.

The "semcheck procedure is designed to be called by any function
with a ptr to a parse tree node as an argument. It will then
determine which sub-module is necessary to check the node.

"R_Initial" presently has the function of initializing the type
table.

Modified : 04/13/87 Output files written to vdisk, "d:" EC

***************************************************************************

Externals

***************************************************************************

#include (semcheck.hpp)

extern void c_startup () ;
    c_ending () ;
    user () ;
    user_end () ;
    close () ;
    set_base () ;
    move_cursor () ;

extern unsigned ;
    alert () ;
    base () ;

***************************************************************************

Globals

***************************************************************************

unsigned_stack = STACKSIZE.

67
/*************** R_Initial ***************

void
r_initial ()
{extern node types [];

strcpy (types [UNTYPED].name, "untyped");
types [UNTYPED].bytes = NULL;
strcpy (types [BOOLEAN].name, "boolean");
types [BOOLEAN].bytes = BOL_BYTES;
strcpy (types [REAL].name, "real");
types [REAL].bytes = REAL_BYTES;
strcpy (types [INTEGER].name, "integer");
types [INTEGER].bytes = INT_BYTES;
strcpy (types [NATURAL].name, "natural");
types [NATURAL].bytes = NAT_BYTES;
}

/******************** SemChecker *******************/

PHITYPE
semcheck (ptr)
/*
 * Breaks Sem Checker
 */
{nodal ptr;
 (extern PHITYPE tkindef (), trtarrow (),
tfunid (), tid (), tconstant (), tactuallist (), tactuals [];
PHITYPE type;

switch (ptr->name) {
    case (ADD_) :
    case (SUB_) :
    case (MULT_) :
    case (RDIV_) :
    case (IDIV_):
    case (COLON_): type = arithop (ptr);
        break;
    case (POS_):
    case (NEG_): type = tprimary (ptr);
        break;
    case (ORLOG_): type = tor (ptr);
        break;
    case (ANDLOG_): type = tand (ptr);
        break;
    case (NEGLOG_): type = tnegation (ptr);
        break;
    case (KINDEF): tkindef (ptr);
        break;
    case (RTARROW_): type = trtarrow (ptr);
        break;
    case (LETDEF): tletdef (ptr);
        break;
    case (KW_ + WHERE_): type = twhere (ptr);
        break;
    case (AUXAND): tauxand (ptr);
        break;
    case (DATAAUXDEF): tdatauxdef (ptr);
        break;
    case (FUNAUXDEF): type = tfunauxdef (ptr);
        break;
    case (FUNID): type = tfunid (ptr);
        break;
    case (ACTUALLIST): type = tactuals (ptr);
        break;
    case (COMMA_):
APPENDIX G
ROCK COMPILER — SCANNER

* Name: scanner
* Description: ROCK Compiler Scanner
* Source: Written in C
* Purpose: parses and tokenizes the input

```c
PUBLIC MAIN SOFTWARE

int Main(int argc, char **argv)
{
    int i;
    for (i = 1; i < argc; i++)
        printf("%s\n", argv[i]);
    return 0;
}
```

The above code is an example of how the scanner works.

```c
int IfKeyword()
{
    char *Keyword = "if",
        *str;
    str = strdup(argv[1]);
    if (*str == 'i' && *(str + 1) == 'f')
        return 1;
    else
        return 0;
}
```

The function `IfKeyword` checks if the current token is an `if` keyword.

```c
int GetToken()
{
    char *Token = argv[1];
    argv[1] = argv[2];
    argv[2] = argv[3];
    argv[3] = argv[4];
    return strlen(Token) + 1;
}
```

This function gets the next token from the input stream.

```c
int Tokenize()
{
    char *Token = argv[1];
    argv[1] = argv[2];
    argv[2] = argv[3];
    argv[3] = argv[4];
    return strlen(Token) + 1;
}
```

The function `Tokenize` is responsible for tokenizing the input stream.

```c
int Parse()
{
    char *Token = argv[1];
    argv[1] = argv[2];
    argv[2] = argv[3];
    argv[3] = argv[4];
    return strlen(Token) + 1;
}
```

This function parses the input stream and identifies tokens.

```c
int IIF()
{
    char *Token = argv[1];
    argv[1] = argv[2];
    argv[2] = argv[3];
    argv[3] = argv[4];
    return strlen(Token) + 1;
}
```

The function `IIF` is used to identify if the current token is an `if` keyword.

```c
int main()
{
    char *argv[50] = {
        "main", "arg1", "arg2", "arg3", "arg4"
    };
    printf("Argument List: \n%s\n", argv[1]);
    argv[1] = argv[2];
    argv[2] = argv[3];
    argv[3] = argv[4];
    return 0;
}
```

The `main` function is the entry point of the program.
return LAST SEQUENCE;

case 'd':
    \text{return} 
    \begin{align*}
    &\text{get_integer} \quad \text{if} \quad \text{lockAhead} = \text{True} \\
    &\text{return} \quad \text{lockAhead} \quad \text{else if} \quad \text{lockAhead} = \text{False} \\
    &\text{return} \quad \text{lockAhead}
    \end{align*}

    \text{else} \quad \text{lockAhead} = \text{True} \\
    \text{return} \quad \text{lockAhead} \quad \text{else if} \quad \text{lockAhead} = \text{False} \\
    \text{return} \quad \text{lockAhead}

    \text{case 'f':}
    \text{\quad return \quad M100C} \\
    \text{else if \quad lockAhead} = \text{True} \\
    \text{\quad lockAhead} = \text{True} \\
    \text{\quad return \quad lockAhead} \quad \text{else if \quad lockAhead} = \text{False} \\
    \text{\quad return \quad lockAhead}

    \text{case 'h':}
    \text{\quad return \quad M100C} \\
    \text{else if \quad lockAhead} = \text{True} \\
    \text{\quad lockAhead} = \text{True} \\
    \text{\quad return \quad lockAhead} \quad \text{else if \quad lockAhead} = \text{False} \\
    \text{\quad return \quad lockAhead}

    \text{case 's':}
    \text{\quad return \quad M100C} \\
    \text{else if \quad lockAhead} = \text{True} \\
    \text{\quad lockAhead} = \text{True} \\
    \text{\quad return \quad lockAhead} \quad \text{else if \quad lockAhead} = \text{False} \\
    \text{\quad return \quad lockAhead}

    \text{case 't':}
    \text{\quad return \quad M100C} \\
    \text{else if \quad lockAhead} = \text{True} \\
    \text{\quad lockAhead} = \text{True} \\
    \text{\quad return \quad lockAhead} \quad \text{else if \quad lockAhead} = \text{False} \\
    \text{\quad return \quad lockAhead}

    \text{case 'u':}
    \text{\quad return \quad M100C} \\
    \text{else if \quad lockAhead} = \text{True} \\
    \text{\quad lockAhead} = \text{True} \\
    \text{\quad return \quad lockAhead} \quad \text{else if \quad lockAhead} = \text{False} \\
    \text{\quad return \quad lockAhead}

    \text{case 'v':}
    \text{\quad return \quad M100C} \\
    \text{else if \quad lockAhead} = \text{True} \\
    \text{\quad lockAhead} = \text{True} \\
    \text{\quad return \quad lockAhead} \quad \text{else if \quad lockAhead} = \text{False} \\
    \text{\quad return \quad lockAhead}

    \text{\quad lockAhead} = \text{True} \\
    \text{\quad return \quad lockAhead} \quad \text{else if \quad lockAhead} = \text{False} \\
    \text{\quad return \quad lockAhead}

    \text{case 'w':}
    \text{\quad return \quad M100C} \\
    \text{else if \quad lockAhead} = \text{True} \\
    \text{\quad lockAhead} = \text{True} \\
    \text{\quad return \quad lockAhead} \quad \text{else if \quad lockAhead} = \text{False} \\
    \text{\quad return \quad lockAhead}
The function checks to see if the input token is a keyword in the language. If it is, the function returns the numeric value of the keyword. If it isn't, the function returns -1. Performs binary search of keyword array.

**MUST KEEP THIS ARRAY IN ALPHABETICAL ORDER**!
APPENDIX H
ROCK COMPILER — PARSER

PUBLIC DOMAIN SOFTWARE

* Name : parser pt I
* File : parser1.c
* Authors : Maj E.J. COLE / Capt J.E. CONNELL
* Started : 10/20/86
* Archived : 12/11/86
* Modified : 04/23/87 No longer set up to work with file of tokens.

This file contains the following modules for the PHI parser:

- BlockBody()
- LetDefs()
- Defs()
- DefAnd()
- QualExp()
- AuxExp()
- AuxDefs()
- AuxAnd()
- Formals()
- Expression()

Algorithm: The main module calls BlockBody() to start the parse off. BlockBody in turn calls LetDefs() first and then QualExp() looking for a valid program. The remaining modules in Pt's 1-3 are called by these when trying to validate a program. The results from the parse are now kept in an abstract syntax tree for type checking and code generation. Various utility functions are used to build the tree and simplify parsing the grammar.

Modified : 12/26/86 Flattened tree output changed to abstract syntax tree form. JC
01/10/87 Corrections to comply with latest definitions of the language. JC
01/27/87 Error Recovery added and files combined. JC
03/20/87 Token buffer implemented for parser. JC
03/29/87 Changed manner errors are handled - required for integration with back-end.
04/23/87 No longer set up to work with file of tokens.
GetToken is called directly thru FillBuff(). JC

76
/* */
/* must use "long" because buffer */
/* holds addresses */
/* use BUFSIZE = 1 in case "real" */
/* place address of "fake address" */
/* of buffer */

and tokenbuf BUFSIZE-1, *ptr = &tokenbuf BUFSIZE;

FILE *posfile, *errorfile; /* working files */

/**********************************************************

root, Parsed in.

int error; *root = NULL;
extern void pclose(), mov cursor();

*errors = 0;

errorfile = fopen("errors.out", "w");
sprintf(errorfile,"%42s n", "ROCKY ERRORS");
close(errorfile);

#define DEBUG
posfile = fopen("parser.out", "w");

/* Exidy

* ByPass PCH, */
/* MOV cursor(), 
error file "WARNING...." etc. "ex" found */
/* compile hit ready program */
/* read it in out of memory */

* parser */
/* src file */
/* error file */
/* error file */

/* Exidy */

/* Exidy */

/* Exidy */

/**********************************************************

* does a post-order walk of the tree with (root) as its head.
* just prints out the node name in the screen now

77
static int i = 0;

if (root != NULL)
    PostOrder(root->lptr);
PostOrder(root->rptr);

switch (root->name)
    case IDENTIFIER:
    case CONSTANT:
    case LITERAL:
        printf (positive, %d, root->name);
        printf (positive, %d, root->index);
        break;
    default:
        printf (positive, %d, root->name);
        printf (positive, %d, root->index);
        break;

if (i == 0) printf (positive, " " );
    else printf (positive, " " );

******************************************************************************

static int root;
int flag;
int i...
*

*<LETDEFS> := let <DEFS> <BLOCKBODY>
*


******************************************************************************

<LETDEFS> := let <DEFS> <BLOCKBODY>


******************************************************************************


/* found something so need to *
 * check for more def's *
 */

/* any errors have been noted, *
 * so press on *
 */

/* end Defs */

*******************************************************************************/

/* root is a ptr to tree/subtree *
 * currently working with *
 */

<DEFAND> ::= and <DEFS>

Where "and" <DEFS> "need not be present.

Note: This function assumes root is not NULL upon entry */

klass K - AND;

/* found "and" so fix tree */

if root->root->DEFAND,LEFT = ERROR

/* note it, try to fix */

/* end ByPass AND */

/* end DefAnd */

*******************************************************************************/

/* root is a ptr to tree/subtree *
 * currently working with *
 */

<QUALEXP> ::= <EXPRESSION> where <AUXEXP>

Where "where <AUXEXP>" need not be present.

/**
 *
 *
 */

/* errors already reported. *
 * attempt to press on *
 */

if root->root->ERROR

/* looking for where expression *
 * found one, fix tree *
 */

root->root->ptr->ERR = WHERE;

/* need AuxExp following WHERE *
 */

/* end byPass WHERE */

/* AuxExp entered "": scan("%c"); */

exit AuxExp

* default - just return 1... *

end AuxExp(); */

*******************************************************************************/

/* root is a ptr to tree/subtree *
 * currently working with *
 */

<AUXEXP> ::= <AUXDEFS> (where <AUXEXP>) *
int flag;

if(flag = AuxDefs(root);) := TRUE) /* need at least one AUXDEF */
   ErrorHandler(line_no, ERR_1, (long) KW WHERE);
if(ByPass(KW + WHERE)) /* looking for multiple WHERE's */
   MakeNewRoot(root, (KW + WHERE), RIGHT);
   AuxExp(&$(*root)->lptr)); /* find AuxExp following WHERE */
   return(flag); /* default - return result of */
   /* first AuxDefs */
   /* end AuxExp */

   /**************************************************************************
   AUXDefs(root) **root;
   /* root is a ptr to tree/subtree */
   /* currently working with */
   * <AUXDEFS> ::= (<DATAAUXDEF | <FUNAUXDEF>) <AUXAND> */
   *                Where "<AUXAND>" need not be present.
   * 
   * reset *temp;
   * flag;
   * ptr;
   * 
   * ptr = ByPass(IDENTIFIER ());
   * temp = CreateNode(IDENTIFIER ());
   * temp->index = ptr;
   * 
   * ByPass(EQUIV ());
   * root = CreateNode(DATAAUXDEF);
   * root->lptr = temp;
   * ExprExpression(&$(*root)->lptr->rptr) := TRUE; /*
   * ifKWWHERE);
   * if(temp->symbol == TOK_ID) /*
   * IF(temp->symbol == TOK_ID) /*
   * otherwise */
   * IfKWWHERE);
   * if(temp->symbol == TOK_ID) /*
   * IF(temp->symbol == TOK_ID) /*
   * otherwise */
   * IfKWWHERE);
   * if(temp->symbol == TOK_ID) /*
   * IF(temp->symbol == TOK_ID) /*
   * otherwise */
   * IfKWWHERE);
/********

if (!flag
    return(NULL);

else
    return(TRUE);

CHECK:

if ((AuxAnd(root) != TRUE))
    ErrorHandler(line_no, ERR_h,
    (long)KW WHERE);

goto CHECK;

return(flag);


/***************************************************

void AuxAnd(root)

NodeRec **root;

/*

<AUXAND> ::= and <AUXDEFS>
/*

Where "and <AUXDEFS>" need not be present.
/*

Note: This function assumes root is not NULL upon entry
*/

if (ByPass(KW AND))
    MakeNewRoot(root, AUXAND, LEFT);
    if ((AuxDefs(&(*root)->rptr) != TRUE))
        ErrorHandler(line_no, ERR_h,
        (long)KW WHERE);
}

/***************************************************

int Formals(root)

NodeRec **root;

/*

<FORMALS> ::= <ID> | '(' <FORMALS> ',' ID ')' */

NodeRec *temp, *workingroot;

long ptr;

if ((ptr = ByPass(IDENTIFIER)))
    *root = CreateNode(IDENTIFIER);
    (*root)->index = ptr;
    return(TRUE);
Expression(root)  
/* root is a ptr to tree's top */  
/* currently working with */  
/* <EXPRESSION> ::= <CONJUNCTION> ( \ <EXPRESSION>) */  
int flag;

if(!(*flag = Conjunction(root)))  
/* look for Conjunction */  
if(Bypass(ORLOG_))  
"will recursively check for /* found, so fix root for return */  
if(! MakeNewRoot(root, ORLOG_ , LEFT));  
"/ w/o following Exp. */  
/* Just note it, no fix */  
ErrorHandler(line_no, ERR8,
    (long)ORLOG_);
    return(ERROR_);
    /* end recursive search */  
    return(flag);
    /* end Expression() */  
/*******************************************************************************/
PUBLIC DOMAIN SOFTWARE

Name : parser pt 2
File : parser2.c
Authors : Maj. E.J. COLE / Capt. J.E. CONNELL
Started : 10/20/86
Archived : 12/11/86
Modified : 01/27/87 - Error Recovery added. JC

This file contains the following modules for the PHI parser:

- Conjunction()
- Negation()
- Relation()
- Relator()
- SimpliExp()
- AddOp()
- MullOP()
- Term()
- Factor()
- Primary()
- Application()
- Actual()

Algorithm : See parser part 1

Declarations:

```c
#include <stdio.h>
#include <parser.h>

extern int line_no;    /* global var, holds current line */
exem int rmarket;      /* global flag - aids in making */
/* PHI deterministic */

int Conjunction(root)    /* root is a ptr to tree/subtree */
    NodeRec **root;    /* currently working with */

/*
<CONJUNCTION> ::= <NEGATION> (  
|  </CONJUNCTION>)
*/

int flag;

if((flag = Negation(root)) == TRUE)  /* look for Negation part */
    if (ByPass(ANDLOG_))         /* will recursively check for */
        (MakeNewRoot(root,ANDLOG_LEFT);         /* will recursively check for */
        if(Conjunction(((root)->rptr)) == TRUE) /* /\ w/o following Neg. */
            ErrorHandler(line_no,ERR8,
                (long)ANDLOG_);//
            return(ERROR_);        /* /\ w/o following Neg. */
            ErrorHander(line_no,ERR8,
                (long)ANDLOG_);//
            return(ERROR_);        /* /\ w/o following Neg. */
            ErrorHander(line_no,ERR8,
                (long)ANDLOG_);//
            return(ERROR_);        /* /\ w/o following Neg. */
            ErrorHander(line_no,ERR8,
```

85
NEGATION ::= RELATION

<RELATION> ::= <SIMPLEXP> <RELATOR> <SIMPLEXP> :
Where <RELATOR> <SIMPLEXP> need not be present.

if (flag & SIMPLEXP) == 0

return(flag);

else if (flag & RELATOR)

MakeNewRoot(flag, type, ERR);

<SIMPLEXP> (root->expr) == TRUE;

ErrorHandler(1, relno, ERR,

return(ERRERROR);

return(flag);

end RELATION

int

Relator();

<RELATOR> ::= | < | > | <= | >= | in | not in

Note: returns the Relator value vice TRUE if found

int flag;

if ((flag = ByPass(EQ, 1))

else if ((flag = ByPass(NEQ, 1))

else if ((flag = ByPass(LEQ, 1))

else if ((flag = ByPass(GEQ, 1))

else if ((flag = ByPass(KW_\text{IN}_, 1))

else if ((flag = ByPass(KW_\text{NOT IN}_, 1))

else if ((flag = ByPass(KW_\text{LESS}, 1))

else if ((flag = ByPass(KW_\text{GREATER}, 1))

86
17.7 A-

**<SIMPLEXP> ::= <TERM> (<ADDOp><SIMPLEXP>)+**

```c
if(flag=ByPass(ADD)) /* do nothing */
else if(flag=ByPass(SUB)) /* do nothing */
else if(flag=ByPass(COLON)) /* do nothing */
else if(flag=ByPass(CAT)) /* do nothing */

return(flag); /* return result of search */
/* end AddOp */
```

```
**<MULOP> ::= * / % (idiv)**
/*
Returns the MulOp value vice TRUE if found */
```

```c
if(flag=ByPass(MULT)) /* do nothing */
else if(flag=ByPass(RDIV)) /* do nothing */
else if(flag=ByPass(IDIV)) /* do nothing */
```
...
```c
int Application(root) 
    NodeRec **root; 
    /* root is a ptr to tree subtree */
    /* currently working with */
    /* <APPLICATION> ::= (<ACTUAL>)+ */
    int flag; NodeRec *tnode; /* temp pointer to node */
    if(flag = Actual(root)) == TRUE) /* look for an actual */
        if(flag = Application(&tnode)) == TRUE) /* look for an actual list */
            MakeNewRoot(root,ACTUALLIST,LEFT);
            (*root)->rptr = tnode;
            if(**root)->rptr->name = ACTUALLIST) /* fix tree so all Actual's */
                MakeNewRoot((*root)->rptr), ACTUALLIST, LEFT);
                /* end if(Application(th0de) */
            else if(flag == ERROR) /* invalid Actualist */
                ErrorHandler(line_no, ERR_K, NULL); /* note it, no fix */
            else return(TRUE); /* either valid Actualist or */
                /* just a single actual */
                /* return ERROR_or FALSE */
                /* based on first look */
                /* end Application() */
    return(flag);
    /* end recursive search */
    /* end Primary() */
    /*******************************************************************/

int Actual(root) 
    NodeRec **root; /* root is a ptr to tree/subtree */
    /* currently working with */
    /* <ACTUAL> ::= <ID>| file<LITERAL>|<CONDITIONAL>|<BLOCK>| */
        /* <DENOTATION>|<COMPOUND>|<ARGBINDING> */
    long ptr; /* ptr to data struct holding the */
    NodeRec *temp; /* actual value of ID, REAL, etc */
    int flag; /* ptr to temp node in the tree */
    if(ptr = ByPass(IDENTIFIER_))) /* checking for ID */
```
*root = CreateNode(IDENTIFIER_);
  (*root)->index = ptr;

  if (ByPass(LIENRTARROW_))
    MakeNewRoot(root, LIENRTARROW_, LEFT);
    if (Actual((**(root)->rptr)) == TRUE)
      return(TRUE);
    else
      ErrorHandler(line_no, ERR_, (long)LIENRTARROW_);
      return(ERROR_);
  
  return(TRUE);

  if (ByPass(KW_ * FILE_))
    *root = CreateNode(KW_ * FILE_);
    if ((ptr = ByPass(LITERAL_)))
      temp = CreateNode(LITERAL_);
      temp->index = ptr;
      (**root)->rptr = temp;
      return(TRUE);
    else
      ErrorHandler(line_no, ERR_1, NULL);
      return(ERROR_);

  if ((flag = Conditional(root)) == FALSE)
    return(flag);
  if ((flag = Block(root)) == FALSE)
    return(flag);

  if ((flag = Compound(root)) == TRUE)
    if ('ByPass(LIENRTARROW_))
      return(TRUE);
    else
      temp = *root;

      if ('IsFormal(temp))
        ErrorHandler(line_no, ERR_o, NULL);
      (*root)->name = FORMAL;
      MakeNewRoot(root, LIENRTARROW_, LEFT);
      if (Actual((**(root)->rptr)) == TRUE)
        return(TRUE);
      else
        ErrorHandler(line_no, ERR_, (long)LIENRTARROW_);
        return(ERROR_);

    else
      if (flag == ERROR_)
        return(ERROR_);
if (!flag && ArgBind.arg(root)) = FALSE)
    return (flag);

if (!flag && ArgBind.arg(root)) = FALSE)
    return (flag);

return FALSE; /* Default, tried everything else */
/* end Actual() */

 returnType; /* Some return type */
PUBLIC DOMAIN SOFTWARE

* Name : parser pt 3
* File : parser3.c
* Authors : Maj E. J. COLE / Capt J. E. CONNELL
* Started : 10/20/86
* Archived : 12/11/86
* Modified : 01/27/87 - Error Recovery added. JC

This file contains the following modules for the PHI parser:
* Conditional() Arm() Block() Compound()
* Elements() Denotation() ArgBind() Op()
* TypeExp() TypeDom() TypeTerm() TypeFacw()
* TypePrimary() PrimType()

Algorithm : See parser part 1

- Modified : 12/26/86 Flattened tree output changed to abstract syntax tree form. JC
- : 01/10/87 Corrections to comply with latest definitions of the language. JC
- : 01/27/87 Error Recovery added and files combined. JC

```c
#include <stdio.h>
#include <parser.h>

extern int rootnet;       /* global flag - aids to */
extern int line_no;       /* making PHI determine */
                          /* global var. current line */
                          /* number of program */

int Conditional()        /* root is a ptr to tree's root */
NodeRec **root;          /* currently working with */

/* <CONDITIONAL> ::= if <ARM> (elsif<ARM>)* (else<EXPRESSION>)* endif */

NodeRec *temp = NULL, *subroot, *workingptr;       /* ptrs to temp nodes in tree */

if(ByPass(KW_ - IF_))
    if(Arm(*temp) = TRUE)
        ErrorHandler(line_no, ERR_m, (long)IF_);    /* note it, try to fix */
        root = CreateNode(KW_ - IF_);               /* set up root for return */
        (*root) = rp = temp;                        /* attach THEN exp to root */
        workingptr = *root;                        /* move working ptr */

while(ByPass(KW_ - ELSEIF_))
    subroot = CreateNode(KW_ - ELSEIF_);        /* attach ELSEIF to tree */
    workingptr = rp = subroot;                  /* end while ELSEIF */
    if(Arm(*temp) = TRUE)
        ErrorHandler(line_no, ERR_m,             /* note it, try to fix */
                        (long)ELSEIF_);          /* set up root for return */
        subroot = rp = temp;                      /* attach THEN exp to ELSEIF */
        workingptr = rp = workingptr ->rp;        /* move working ptr down subtree */

if(ByPass(KW_ - ELSE_))
```

92
Arm(root); * root is a ptr to tree subtree * current working with * 

/*  
<ARM> ::= <EXPRESSION>then<EXPRESSION> */

if flag;
NodeRec **temp = NULL; /* temp ptr to a ptree in 'tree' */

if(flag = Expression(temp)) = TRUE)  
  EatEm(KW_"THEN"); /* if an error try to recovery */
  look for THEN,ELSE,ELSEIF,END/*

if (ByPass(KW_"THEN",))  
  root = CreateNode(KW_"THEN"); /* sets root for return errors */
  (*root) = lptr = temp;
  if (Expression(temp) == TRUE)  
    (*root) = rptr = temp;  
  else  
    ErrorHandler(line_no,ERR_r,  
      (long)THEN"); /* report it and try to press in */
  
else  
  ErrorHandler(line_no,ERR_f,  
      (long)KW_"THEN");  
return(flag);  
/* end Arm() */

/** *******************************************************************************/

Block(root); /* root is a ptr to tree subtree */
NodeRec **root; /* currently working with */

/*  
<BLOCK> ::= begin <BLOCKBODY> end */

if (ByPass(KW_"BEGIN",))  
  root = CreateNode(KW_"BEGIN");  
  /* sets root for return errors */
  have already been reported */

if (BlockBody( IF(*root) -> lptr)) = TRUE)  
  look for BLOCKBODY */

93
where <ELEMENTS> may be empty

```c
1. Pass LPAR:
   Element(root);
   ErrorHandler(line_no, ERR_PAR);
   ByPass(LPAR);
   ErrorHandler(line_no, ERR_PAR);

   if (root == NULL)
     root = CreateNode(EMPTYCOMPOUND);
   else if ((root)->name == COMMA)
     (root)->name = ELLIST;
   return(TRUE);

2. Pass RTSQUIG)
   Elements(root);
   ErrorHandler(line_no, ERR_RT);
   ByPass(RTSQUIG);

   if (root == NULL)
     root = CreateNode(EMPTYCOMPOUND);
   else if ((root)->name == COMMA)
     (root)->name = ELLIST;
   return(TRUE);

3. Pass ST_SEQUENCE_) 
   Elements(root);
   if (!ByPass(END_SEQUENCE_))
     ErrorHandler(line_no, ERR_RT);
     (long)END_SEQUENCE_);

   if (root == NULL)
     root = CreateNode(EMPTYSEQUENCE);
   else
     MakeNewRoot(root, SEQUENCE, RIGHT);
   return(TRUE);

   return(FALSE);
```
int ElementRec (root)  
NodeRec *root;  
/ * root is a ptr to tree */  
/ * currently working with */
/ *
<ELEMENTS> ::= <QUALEXP> (,<QUALEXP>) *
/ *
int flag;
if (flag = QualExp(root)) == ERRGR_  
EatExp(COMMA_);  
while (ByPass(COMMA_))  
/* errors already set now */  
/* recursively look for root */  
/* qualexp */
MakeNewRoot(root,COMMA_,LEFT_);  
if (Elements(&(*root)->rptr)) == TRUE)  
ErrorHandler(line_no,ERR_p,  
(Long)COMMA_);  
if ((*root)->rptr->name == COMMA_)  
MakeNewRoot(&(*root)->rptr),  
COMMA_,LEFT_);  
/* fix tree so far */  
/* hang to the liii */  
/* end while ByPass */
return(flag);
/ * end Elements */
/ ***********************************************
int Denotation(root)  
NodeRec *root;  
/ * root is a ptr to tree */  
/ * currently working with */
/ *
<DENOTATION> ::= <LITERAL> | <CONSTANT> : <FORMALS> ...
* where LITERAL is quoted(') string of zero or more characters
* where CONSTANT is an integer or decimal number
* NOTE: <FORMALS> ::= <ACTUAL> was already checked to be ...
Long ptr;
if (ptr = ByPass(LITERAL_))  
  *root = CreateNode(LITERAL_);  
  (*root) ->index = ptr;  
  return(TRUE);
if (ByPass(EMPTY_LIT_))  
  *root = CreateNode(LITERAL_);  
  (*root) ->index = NULL;  
  return(TRUE);
if (ptr = ByPass CONSTANT)  
  *root = CreateNode(CONSTANT)  
  (*root) ->index = ptr;  
  return(TRUE);  
/ */
/**<ARGBINDING>**:=('[<OP><QUALEXP>|<QUALEXP><OP>|<OP>])'*/

int special_case;
NodeRec *temp = NULL; /* temp ptr to node in tree */
extern int argbind; /* global flag needed to make */
PHI deterministic; /* */
if (ByPass(RTBRAKET))
argbin = TRUE; /* set global flag, needed to */
PHI deterministic; /* */
specialcase = (Ball(ADD,1) | Ball(SUB,1));

def ERROR
printf("special case \%d argbind = \%d\n", specialcase, argbind);
endf

if (Op(root))
/* begin Op comes first */
if (ByPass(RTBRAKET))
/* looking for <Op> */
argbind = FALSE; /* reset global flag */
return(TRUE);

MakeNewRoot(root, ARGBINDOP, LEFT);
return(TRUE); /* had <Op> */

MakeNewRoot (root, ARGLEADOP, LEFT);
if ((Ball(ADD,1) | Ball(SUB,1))
specialcase = FALSE;

if ((QUALEXP(!(*root)->rptr) ==TRUE)
if (ByPass(RTBRAKET))
argbind = FALSE; return(TRUE); /* reset global flag */
else
if (specialcase && Op(temp)
44 ByPass(RTBRAKET))
if (*root)-lptr->rptr(*root)->rptr;
(*root) = *rptr = temp;
if (*root)-lptr->name == ADD_
if (*root)-lptr->name = POS_
if (*root)-lptr->name = NEG_
if (*root) = name = ARGTRAILOP;
argbind = FALSE; /* <Op> came last as a "," */
/* reset global flag */
return(TRUE); /* */

argbind = FALSE;
ErrorHandler(line_no, ERR_q, NULL);
return(ERROR_); /* */

if ((QUALEXP(root)) == FALSE)
MakeNewRoot (root, ARGTRAILOP, LEFT);
argbind = FALSE;
if (Op(*root)->rptr)
44 ByPass(RTBRAKET))
return(TRUE); /* see if can continue */
ErrorHandler(line_no, ERR_q, NULL);
return(ERROR_); /* */

return(FALSE); /* */

/*************************/
Op(root)                   /* root is a ptr to tree/subtree */
    NodeRec **root;       /* currently working with */
    /*<OP> ::= , | ! | <RELATOR> | <ADDOP> | <MULOP> */

int flag;

if(flag == ByPass(COMMA_))
    *root = CreateNode(COMMA_);
else if(flag == ByPass(SUBSCRIPT_))
    *root = CreateNode(SUBSCRIPT_);
else if(flag == Relator())
    *root = CreateNode(flag);
else if(flag == AddOp())
    *root = CreateNode(flag);
else if(flag == MulOp())
    *root = CreateNode(flag);
return(flag);
    /* end Op
    ***************************************************/

int TypeExp(root)        /* root is a ptr to tree/subtree */
    NodeRec **root;       /* currently working with */
    /* <TYPEEXP> ::= <TYPEDOM> ( -> <TYPEEXP> ) */

NodeRec *newroot;        /* temp ptr to nodes in the tree */
int flag;

if(((flag = TypeDom(root)) == TRUE)
    if (ByPass(RTARROW_)) /* will recursively search for */
    { newroot = CreateNode(RTARROW_); /* more TYPEEXP's */
      *root = newroot;
      if ((TypeExp((&(*root)->right)) != TRUE)
        ErrorHandler(line_no, ERR9, (long) RTARROW_);
        return(ERROR_);
    }
    return(flag);              /* end recursive search */
/* end TypeExp */
    ***************************************************/

int TypeDom(root)        /* root is a ptr to tree/subtree */
    NodeRec **root;       /* currently working with */
    /* <TYPEDOM> ::= <TYPETERM> (+ <TYPEDOM>) */

NodeRec *newroot;        /* temp ptr to nodes in the tree */
int flag;

if(((flag = TypeTerm(root)) == TRUE)
    if (ByPass(ADD_)) /* will recursively search for */
    { newroot = CreateNode(TYPEPLUS);
      newroot ->lptr = *root;
      *root = newroot;
      if ((TypeDom(&(*root)->right)) != TRUE)
        ErrorHandler(line_no, ERR9, (long) RTARROW_);
        return(ERROR_);
    }
    return(flag);
/* end TypeDom */
    ***************************************************/

97
newroot;
if (TypeDom(&(('*root')->rptr)) != TRUE)
    ErrorHandler(line_no, ERR9, (long)ADD_);
    return(ERROR_);
}
return(flag);
/* end TypeDom() */

/****************** end recursive search **/
int
TypeTerm(root)
    /* root is a ptr to tree/subtree */
    /* currently working with */
/
/*
<TYPETERM> ::= <TYPEFAC>(('*' <TYPETERM>)) */
/
NodeRec *newroot;
/* temp ptr to nodes in the tree */
int flag;
if((flag = TypeFac(root)) == TRUE)
    if (ByPass(MULT_)) /* will recursively search for */
        /* more TYPETERMS's */
        /* fix root for return */
        newroot = CreateNode(TYPETIMESI;
        newroot ->lptr = *root;
        *root = newroot;
    if(TypeTerm(&(('*root')->rptr)) != TRUE)
        /* end recursive search */
        return(flag);
    /* end TypeTerm() */
/* end recursive search */
int
TypeFac(root)
    /* root is a ptr to tree/subtree */
    /* currently working with */
/
/*
<TYPEFAC> ::= <TYPEPRIMARY>|<TYPEFAC> | <TYPEPRIMARY> */
/
/*
<ID> "<<" <TYPEEXP> (,<TYPEEXP>)* '"' >' <ACTUAL> */
/
/* Where <<TYPEEXP(,TYPEEXP,...)>> and/or <ACTUAL> */
/
/* need not be present */
/
NodeRec *newroot;
/* temp ptr to nodes in the tree */
int flag;
long ptr;
if(ptr = ByPass(IDENTIFIER_)) /* root = CreateNode(IDENTIFIER_); */
    (*root) ->index = ptr;
    if(ByPass(ST_SEQUENCE_) & ByPass(ST_SEQUENCE_))
        /* end bypass << */
        goto CHECK;
    /* end if ID */
    if((flag = TypePrimary(root)) == TRUE)
        goto CHECK;
    return(flag); /* return either ERROR or FALSE */
CHECK: if(ByPass(STAR_))
    { newroot = CreateNode(STAR_);
      newroot ->lptr = (*root);
      *root = newroot;
    } /* end if STAR */

    return(TRUE); /* made it this far, all OK */
} /* end TypeFac() */
/* ***********************************************/

int
TypePrimary(root) NodeRec **root; /* root is a ptr to tree/subtree */
/* currently working with */
/* */
/* NOTE: ID already checked in TYPEFAC() */
/* */
if(ByPass(LTPAREN_))
    { if(TypeExp(root) !=TRUE)
      ErrorHandler(line_no,ERR9, /* note it, no fix */
        (Long)LTPAREN_);
      if(ByPass(RTPAREN_))
        return(TRUE);
      else
        ErrorHandler(line_no,ERR_f, /* note it, no fix */
          (Long)RTPAREN_);
        return(ERROR_);
    }
    /* end ByPass '(' */
}
if(PrimType(root))
    return(TRUE);
return(FALSE); /* default */
} /* end TypePrimary() */
/* ***********************************************/

int
PrimType(root) NodeRec **root; /* root is a ptr to tree/subtree */
/* currently working with */
/* */
/* <PRIMTYPE> ::= real | integer | natural | boolean | trivial | type */
{
    if(ByPass(REAL_))
        { *root = CreateNode(REAL_);
          return(TRUE); /* end if REAL */
        }
    if(ByPass(INTEGER_))
        { *root = CreateNode(INTEGER_);
          return(TRUE); /* end if INTEGER */
        }
    if(ByPass(NATURAL_))
        { *root = CreateNode(NATURAL_);
          return(TRUE); /* end if NATURAL */
        }
    if(ByPass(BOOLEAN_))
        { *root = CreateNode(BOOLEAN_);
return(TRUE);

if(ByPass(TRIVIAL_))
{
    *root = CreateNode(TRIVIAL_);
    return(TRUE);
}  /* end if TRIVIAL */

if(ByPass(KW_ TYPE_))
{
    *root = CreateNode(KW_ TYPE_);
    return(TRUE);
}  /* end if TYPE */

return(FALSE);  /* default - none of the above */  /* end PrimType() */

*************************************************************************/


PUBLIC DOMAIN SOFTWARE

Name: Parser Utilities
File: parsr_util.c
Authors: Maj E.J. COLE / Capt J.E. CONNELL
Started: 01/26/87
Archived: 03/03/87
Modified: 04/23/87 FillBuffer() now calls GetToken() direct.

This file contains the utility modules for the parser:
- CreateNode()
- MakeNewRoot()
- ByPass()
- FillBuff()
- IsFormal()
- IBall()
- NodeName()
- EnterName()
- FindName()

Modified: 03/20/87 - Buffer Handling routines added - JC
04/23/87 - FillBuffer() calls GetToken() direct vice
working with intermediate file of tokens.
- EnterName() and FindName() added to place
IDs, LITERALS, and CONSTANTS into the name
table. JC

#include <stdio.h>
#include <parser.h>

extern int line_no; /* global var, holds line no */
extern FILE *pinfile; /* of source prog */
extern FILE *w outfile; /* global working file */

char token[MAXLINE] = "x"; /* Init token[0] to value other 
than NULL. Token[0] holds the 
length of the string. */
NameRec *name_table; TABLESIZE- 1; /* add 1 because 0 is unused */
*EnterName();

UTILITY

/*
NodeRec *
CreateNode(op)
NodeType op; /* operator type of node */

/* Creates a tree node and returns the pointer (temp) to this nod */
/* Accepts node type (op), an integer, and inserts it into the node. */

NodeRec *temp;

temp = CALLOC(1,NodeRec); /* create a node */
temp -> name = op;
temp -> ln = line_no;
temp -> lptr = (temp -> rptr) = NULL;
return(temp); /* end CreateNode() */
NodeRec **root; /* old root of subtree */
int type, side; /* will turn into new root */
  /* (type) is type of new root */
  /* (side) is side to att old root */

/* Creates a new working root for subtree. */
/* Old root is attached to lt/rt based on value of (side) */

NodeRec *newroot;

newroot = CreateNode(type);
(side == LEFT) ?
  (newroot ->lptr = *root) : (newroot ->rptr = *root);
*root = newroot;

/************************************************************************/

void FillBuff(start)
long *start; /* which slot in the buffer */
/* array to start the filling */

/* Requires the buffer array and buffer ptr to be previously defined. */
/* Fills the buffer with tokens by calling GetToken(). Buffer filled */
/* until 1) end of user prog reached or 2) end of the array reached */
/* If the token is a literal, id, or constant then EnterName() is */
/* called to enter it into the nametable. */
/* Lastly, resets the buffer ptr to tokenbuff[0]. */

extern long tokenbuff[ ]; *ptr;
int token_num;
NameRec *nptr; /* identifies a token type */
/* ptr to structure of NameRec */

ptr = start; /* init ptr to travel thru buff */

do
  token_num = GetToken(token);
  *ptr = token_num;
  -*ptr;

  switch (token_num)
  
  case LITERAL : /* insert length of sting */
    token[0] = strlen(token);
    if((nptr=EnterName(token)))
      *ptr = (long)nptr;
      -*ptr;
    else ErrorHandler(NULL, ERR7, NULL); /* HANDLE MEMORY OVERFLOW */
    break;
  case CONSTANT : /* do nothing */
    break;
  case IDENTIFIER : /* end switch */
  case EOF : /* end case */
    break;
  default : /* do nothing */
  /* end switch */
  while((token_num != EOF) && (ptr < &tokenbuff[BUFSIZE]));

  ptr = &tokenbuff[0]; /* reset the buffer ptr */
  /* end FillBuff() */

************************************************************************/
long
ByPass(tgt)
    int tgt;

    /* Checks to see if the next token in the buffer matches the target. */
    /* If so, then returns the token no. and increments the buffer */
    /* pointer */
    
extern long tokenbuff[], *ptr;

    if(ptr >= &tokenbuff[BUFSIZE])
        FillBuff(&tokenbuff[0]);
    /* see if at end of buffer */
        while(*ptr == EOLN_)
    /* increment counter & skip */
    ++ptr;
    ++line_no;
    if(ptr >= &tokenbuff[BUFSIZE])
        FillBuff(&tokenbuff[0]);
    /* see if at end of buff */
    /* end while */
    
    if (*ptr != tgt)
        return(FALSE);
    /* otherwise, it was found */
    
    if(ptr == &tokenbuff[BUFSIZE])
        FillBuff(&tokenbuff[0]);
    /* if at end of buffer */
    
switch (tgt)
    case LITERAL  :
        case IDENTIFIER :
            case CONSTANT :
                /* return ptr to struct */
                return(*((ptr++));
                /* holding the token */
        default:
            return(tgt);
            /* end switch */
            /* end ByPass() */

    /**************************************************************************

    /***********************isFormal*********************************************************

int
    isFormal(root)
        NodeRec *root;
    /* root is ptr to subtree */
    /* currently working with */

    /* Required to make the language deterministic. Compound() returned */
    /* TRUE and "|->" was subsequently found. Formal is a proper subset of */
    /* the compounds so need to insure no errors in the formals. */
    /* */
    /* Performs a preorder search of the subtree. NOTE: assumes that root */
    /* initially points to a non-null compound list. */

 ifdef DEBUG
    printf("isformal entered, root->name = %d\n",root->name);
    if (root == NULL) printf("root is null\n");
    endif
    if(root == NULL)
        return(TRUE);
    
    if(root->name==COMMA  : root->name=IDENTIFIER_
        : root->name==ELLIST)

103
if((IsFormal(root->iptr))
   " (IsFormal(root->rptr)))
   return(TRUE);

return(FALSE);
} /* end Isformal */

int Isall(tgt, index)
int tgt, index;
/* Checks to see if the (index)th token in the buffer matches the */
/* target. If it does returns TRUE else FALSE. Does not increment */
/* the buffer pointer. Checks for full buffer implemented in this */
/* manner to allow for future flexibility. Could have used simple */
/* heuristic of: */
/* if(ptr + (3*index) > &tokenbuff[BUFSIZE]) RefilBuffer; */
/* at the expense of generality */
extern long tokenbuff[], *ptr;
long *tptr;

if(ptr == &tokenbuff[BUFSIZE])
   FillBuff(tokenbuff[0]);
/* see if at end of buff if */
/* start over if had to refill */
DO_AGAIN:
   tptr = ptr;
   while(*tptr == EOLN_)
      /* increment tptr & skip EOLNs */
      ++tptr;
      if(tptr == &tokenbuff[BUFSIZE])
         goto REFIL;
   /* see if at end of buff */
   /* need to refill buffer and */
   /* then start over */
   /* only enter for loop if need to */
   for(; index > 1; --index)
      switch (*tptr)
      {
      case IDENTIFIER:
         case CONSTANT-:
         case LITERAL:
            tptr -= 2; break;
      case EOLN:
         while(*tptr == EOLN_)
            /* increment counter & skip */
            ++tptr;
         if(*tptr == &tokenbuff[BUFSIZE])
            goto REFIL;
      default:
         /* end switch */
      ++tptr;
      if(tptr == &tokenbuff[BUFSIZE])
         goto REFIL;
      /* check if will overflow buff */
      if (*tptr != tgt) return(FALSE);
      else return(TRUE);
   /* take what's left in buffer, */
   /* put at beginning, now refill */
   /* test of buffer */
   for(*tptr = &tokenbuff[0];
      ptr < &tokenbuff[BUFSIZE]; ptr++, tptr++)
      /* refill buffer from current */
      /* posit to end */
goto DO_AGAIN;  
/* refilled buffer, so start over */  
/* end :Ball() */  
/*******************************************************************/

char *
NodeName(ptr)
NodeRec *ptr;
  
/* Accepts a ptr to a structure of NodeRec. Dereferences this node to get a ptr to a structure of NameRec which hold the string containing the name of the value in NodeRec. Returns the name to calling routine */
  
NameRec *temp;  
  
  temp = (NameRec *)(ptr->index);  
  return(temp->name + 1);  
  
/*******************************************************************/
APPENDIX I
ROCK COMPILER — ERROR HANDLER

/* PUBLIC DOMAIN SOFTWARE */

* Name: Error Handler
* File: errors.c
* Authors: Maj E.J. COLE / Capt J.E. CONNELL
* Started: 01/20/87
* Archived: 04/07/87
* Modified:

This file contains the execution modules for error recovery.
ErrorHandler(), EatEm()

Algorithm: ErrorHandler() is called by other modules in the
compiler. It insures the error count is updated and
the error is written to the error file. If required,
ErrorHandler() calls EatEm() to gobble tokens to get to
a known point in the parse. Used during error
recovery. After MAXERRORS number of errors simply
returns to calling routine.

* NOTE: 'errorfile' must have been initially created before
* ErrorHandler() is first called - don't want to append
to last times errors!

* Modified:

#include <stdio.h>
#include <scanner.h>
#include <errors.h>

extern FILE *errorfile; /* working file */
int num_errors = 0; /* running tally of # errors */
char *errors[] = {
  /* incomplete '->'' ",
  " "RESERVED FOR FUTURE USE",
  "'\" without following '(', logical OR is '\"',
  "$' without following 'R', 'N', 'Z', 'B', or 'I',
  "invalid numeric constant ==> " ,
  "literal without ending - " ,
  "unidentified char in input program ==> " ,
  "MEMORY OVERFLOW DURING COMPILATION", 
  "error in statement following ==> " ,
  "error in type definition following ==> " ,
  "unable to complete definition of blockbody after keyword LET", 
  "missing or misplaced ';' after definition", 
  "valid qualexpr/exp not found in the def's xdef".
};
execl("rock.exe", "rock.exe", NULL);

fprintf(errorfile, "line %d : %s
", line_no, errors[err_no]);

switch (err_no) {
    case ERR4:
        case ERR5: fprintf(errorfile, "%s
", (char *)str_num); break;
        case ERR6: fprintf(errorfile, "%s
", (char *)str_num); break;
        case ERR8: switch(str_num) {
            case LEQ_: fprintf(errorfile, "<=\n"); break;
            case NEQ_: fprintf(errorfile, "<>\n"); break;
            case GEQ_: fprintf(errorfile, ">=\n"); break;
            case ADD_: fprintf(errorfile, "+
"); break;
            case SUB_: fprintf(errorfile, "-
"); break;
            case EQ_: fprintf(errorfile, "=
"); break;
            case MULT_: fprintf(errorfile, "*
"); break;
            case IDIV_: fprintf(errorfile, "/\n"); break;
            case SUBSCRIPT_: fprintf(errorfile, "!
"); break;
            case COLON_: fprintf(errorfile, ":\n"); break;
            case ORLOG_: fprintf(errorfile, "\|
"); break;
            case ANDLOG_: fprintf(errorfile, "&\|
"); break;
            case NEGLOG: fprintf(errorfile, "-\n"); break;
            case CAT_: fprintf(errorfile, "\n"); break;
            case LERTARROW : fprintf(errorfile, "->\n"); break;
            case (KW_GREATER ) : fprintf(errorfile, "GREATER\n"); break;
            case (KW_+IN_-) : fprintf(errorfile, "IN\n"); break;
            case (KW_LESS_): fprintf(errorfile, "LESS\n"); break;
            case (KW_NOTIN_): fprintf(errorfile, "NOTIN\n"); break;
            default:
                fprintf(errorfile, "UNDEFINED error\n");
        } /* end switch case ERR8 */ break;

    case ERR9: switch(str_num) {
        case ADD_ : fprintf(errorfile, "+
"); break;
        case MULT_ : fprintf(errorfile, "*
"); break;
        case RTARROW : fprintf(errorfile, "->\n"); break;
        case RTPAREN : fprintf(errorfile, "()\n"); break;
        default:
            fprintf(errorfile, "UNDEFINED error\n");
        } /* end switch case ERR9 */ break;

    case ERR_{: switch(str_num) {
        case KW_+AND_:
            fprintf(errorfile, "==\n"); break;
        case RTFAREN : fprintf(errorfile, ")\n"); str_num(NULL); break;
        case RTSQUIG : fprintf(errorfile, "(\n"); str_num(NULL); break;
        case END_SEQUENCE:
            fprintf(errorfile, "(\n"); str_num(NULL); break;
    } /* don't want to go to EatE-

108
case KW_END:
    printf(errorfile,"KEYWORD END\n\n");
    str_num = KW_END; break; /* set up for calls to eatem() */
case KW_THEN:
    printf(errorfile,"KEYWORD THEN\n\n");
    break;

default:
    printf(errorfile,"UNDEFINED error\n\n");
    /* end switch case ERR */
    break;

case ERR_TH: switch(str_num)
    case IF:
        printf(errorfile,"IF\n\n"); break;
    case ELSIF:
        printf(errorfile,"ELSIF\n\n"); break;
    case ELSE:
        printf(errorfile,"ELSE\n\n"); break;
    case THEN:
        printf(errorfile,"THEN\n\n"); break;
    case BEGIN:
        printf(errorfile,"BEGIN\n\n"); break;
    default:
        printf(errorfile,"UNDEFINED error\n\n");
        /* end switch case ERR_TH */
        break;

str_num = KW_THEN; /* set str_num up to be passed */
break;

default: printf(errorfile,"\n\n"); /* to Eatem() */
end switch;

fclose(errorfile);

if ((err_no = ERROR) &&
    (err_no < ERROR) &&
    (str_num != NULL))
    Eatem((int)str_num);

/**************************************************************************

void
Eatem(tgt);
int tgt;

/* Increments token buffer pointer until tgt token is found. */
/* Use in error recovery to reach a known point in the program. */
extern long tokernumoff, *ptr;
extern int line_no;

#define DEBUG
printf("eatem entered, tgt = %d", tgt);
/*end:

while (!ptr == EOF)
    switch (tgt)
        case EOLN:
            *ptr = ++line_no; break;
        case SEMI:
            if (*ptr == SEMI) (*ptr++) = 1;
return;
--ptr; break;

case EQUIV :
  switch ((int)*ptr)
  case EQUIV :
    case SEMI :
    case KW _AND :
    case KW _AND : return;
    default : ++ptr;
    break;
    /* end switch case EQUIV */
  case KW _WHERE :
    switch ((int)*ptr)
    case KW _WHERE :
    case KW _AND :
    case KW _AND : return;
    default : ++ptr;
    break;
    /* end switch case WHERE */
  case KW _AND :
    switch ((int)*ptr)
    case KW _AND :
    case KW _AND : return;
    default : --ptr;
    break;
    /* end switch case AND */
  case RTPAREN :
    switch ((int)*ptr)
    case RTPAREN :
    case LTPAREN :
    case RTPAREN : return;
    default : --ptr;
    break;
    /* end switch case RTPAREN */
  case KW _IF :
  case KW _ELSE :
  case KW _ELSE :
  case KW _THEN :
    switch((int)*ptr)
    case KW _ELSIF :
    case KW _ELSE :
    case KW _ENDIF :
    case KW _THEN : return;
    --ptr; break;
    /* end switch case THEN, etc */
  case COMMA :
    switch ((int)*ptr)
    case COMMA :
    case RTPAREN :
    case RTPAREN :
    case LTPAREN :
    case RTPAREN :
    case LTPAREN :
    case RTPAREN :
    case STSEQUENCE :
    case ENDSEQUENCE :
    case SEMI :
    case KW _LET :
    case KW _WHERE :
    case KW _AND :
    case KW _AND : return;
    default : --ptr;

case KW_END_
    : 
    case KW_BEGIN_
        : switch ((int)*ptr)
            case KW_END_
            case KW_LET_
            case KW_WHERE_
            case KW_AND_
            case COMMA_
            case RTPAREN_
            case RTSQUIG_
            case END_SEQUENCE_
                case SEMI_
                    return;
                default
            **ptr;
        break;
    return; /* end switch case BEGIN/END */
    default:
        return; /* end switch */
    } /* end while */
    } /* end EatEm() */
/***************************************************************************/
APPENDIX J
ROCK COMPILER — SEMANTIC CHECKER

/* This file contains the following modules for the PHI parser: */
/*
 Hnumconvert Numconvert */
/*
 Algorithm:
 This module contains procedures for type conversion. If the rt child of a node may be converted to the lt type but the con-
 verse is not true, "Hnumconvert" is called. If either side may be converted, "numberconvert" is called */
/*
 Modified: */

/*************************************************************** **
****************************** Externals ***************************/
#include <semcheck.h>

extern void terror (

PHITYPE
hnumconvert (ltype, rtype, ptr) /* Type conversions for the right side of the tree only */

PHITYPE ltype, rtype;

nodal ptr;

extern void c_ztor();

if ((ltype == BOOLEAN) && (rtype == BOOLEAN)) /* No type conversion needed */
return (BOOLEAN);

switch (ltype) /* Predicate actions on type of lt */

case (REAL) : switch (rtype) { /* side of node */

    case (REAL) : return (REAL); /* Matching types; no conv req */
    case (INTEGER) :
    case (NATURAL) :
        c_ztor (); /* Generate code for conversion */
    default :
terror (ERR_aa, ptr->ln); /* No appropriate match: error */
return (REAL); /* Ret real so semantic check cont. */

case (INTEGER) : switch (rtype) {
    case (INTEGER) :
    case (NATURAL) : return (rtype);
    default :
    terror (ERR_cc, ptr->ln);
    return (INTEGER); }

case (NATURAL) :
    if (rtype == NATURAL) return (rtype);
    else {
    terror (ERR_bb, ptr->ln);
    return (NATURAL);
    }
    default : terror (ERR_aa, ptr->ln);
    return (NATURAL);
}

/************************************************************* Numconvert **********************************/
PHITYPE
numconvert (ptr) /* Do number conversions for */
    nodal ptr;
    (PHITYPE ltype, rtype);
    extern PHITYPE semcheck ();
    extern void c_ztor ();

    ltype = semcheck (ptr->ptr);
    /* Get left type */

    if (ptr->rptr->name == (KW_ENDIF)) /* Special case of "if" sequence */
    return (ltype);
    rtype = semcheck (ptr->rptr);
    /* Get right type */

    if ((ltype == BOOLEAN) && (rtype == BOOLEAN)) /* No conversion necessary */
    return (BOOLEAN);

    switch (ltype) {
    case (REAL) : switch (rtype) {
    case (REAL) : return (REAL); /* Types are same: no action needed */
    case (INTEGER) :
    case (NATURAL) :
    c_ztor ();
    return (REAL); /* Generate code for int to real */
    default :
    terror (ERR_aa, ptr->rptr->ln);
    return (REAL); /* No conversion possible */
    }

    case (NATURAL) : switch (rtype) {
    case (REAL) :
    c_ztor (); /* Convert left side */
    return (REAL);
    case (INTEGER) :
    c_ztor (); /* No conversion necessary */
    case (NATURAL) :
    return (NATURAL);
    default :
    terror (ERR_aa, ptr->rptr->ln);
    return (NATURAL); /* No conversion necessary */
    }
case (INTEGER) : switch (ctype) ;
  case (REAL) : 
    c_ztor ();
    return (REAL);
  case (INTEGER) :
  case (NATURAL) :
    return (INTEGER);
  default :
    terror (ERR_aa, ptr->rptr->ln);
    return (NATURAL);
}

default :
  terror (ERR_aa, ptr->lptr->ln);
  /* Types are not numeric
   */
  return (NATURAL);

 /* Convert left side
   */
PUBLIC DOMAIN SOFTWARE

Name: Semcheck Module 1
File: Seml.c
Authors: Maj E.J. COLE / Capt J.E. CONNELL
Started: 01/02/87
Archived: 01/10/87
Modified:

This file contains the following modules for the PHI parser:

- Tletdef
- Trtarrow
- Tkindef
- Twhere
- Tdataauxdef
- Tauxand
- Taridcheck
- Tauxand
- Ttypetimes

Algorithm:
This module contains scoping procedures (Twhere and Tauxand) definition procedures (trtarrow, tkindef, ttypetimes) and the data definition procedure.

Modified:

******************************************************************************** Externals ********************************************************************************

/*include <semcheck.h>
*/
#include <string.h>
extern int typeptr; /* For "strcpy"
extern node types ();
extern void terror ();

tnode *fhead = NULL;

/*************************************************************************** Tletdef ***************************************************************************/

void
tletdef (ptr)
{
    node ptr;
    semcheck (ptr->lptr);
    semcheck (ptr->rptr);
}

/*************************************************************************** Trtarrow ***************************************************************************/

PHITYPE
trtarrow (ptr)
{
    node ptr;
    PHITYPE type, rtype;
    extenr void putform ();

    type = semcheck (ptr->lptr);
    rtype = semcheck (ptr->rptr);

    if ((ptr->lptr->name == TYPETIMES) ||
        (ptr->rptr->name == TYPETIMES))
        putform (type);

    return (rtype);
}
void tkindef (ptr)
    nodal ptr;
    (extern defptr defhead;
    extern void putdef ();
    PHITYPE rtype;

    rtype = semcheck (ptr->rptr);
    putdef (rtype, ptr->lptr);
    defhead->fptr = fhead;
    fhead = NULL;

/** ********************************** Tkindef **********************************/

/** ********************************** Twhere **********************************/

/** ********************************** TDatauxdef **********************************/
while (*holder - count) != NULL) {  /* Push piano through the door  */     /* to copy strings  */
    (ptr->lptr->label | count) = (*holder - count));  /*count;  
    --count;  */

    c_store_code ("ret
'");  // Generate code to end procedure  
    c_store_code (name);  /* CANNOT USE C_END_PROC () HERE;  */
    c_store_code (":n\n");  /* NO SCOPE CHANGE;  */

*/ ************ And_Check *******************/

void  
and_check (mark, ptr, mark_and)  
{  
    varptr mark;  
    and_ptr *mark_and, ptr;  
    extern varptr varhead;  
    extern int buff_ptr;  
    extern char *code_buffer;  
    int buff_holder;  
    varptr v_ptr = Varhead;

    if (ptr != NULL) {  
        and_check (mark, ptr->link, mark_and);  
        do {  
            if (v_ptr->nptr->index==ptr->ptr->index)  
                buff_holder = buff_ptr;  
                buff_ptr = ptr->buffptr;  
                c_call_proc (v_ptr->nptr->label);  
                buff_ptr = buff_holder;  

                if (*mark_and == ptr)  
                    *mark_and = ptr->link;  
                    del_and (ptr);  
                    break;  
            
            if (v_ptr == mark) break;  
            v_ptr = v_ptr->link;  
        } while (TRUE);  
        if (v_ptr == mark) break;  
        v_ptr = v_ptr->link;  
    } while (TRUE);  
    
}  

}  

/* ************ Tauxand *******************/

void  
tauxand (ptr)  
{  
    varptr mark;  
    extern FLAG and_flag;  
    extern and_ptr and_head;  
    int save_and;  
    varptr mark;  
    and_ptr ptr, mark_and = and_head;  
    save_and = and_flag;  
    and_flag = TRUE;  

}  


semcheck (ptr->lptr);
mark = varhead;
semcheck (ptr->rptr);

and_check (mark, and_head, &mark_and);
and_flag = save_and;

/* Semantic Check */

/** Check all new funcn & data defs */

/* Restore and flag */

tptr = and_head;

while (tptr != NULL)

/* Traverse list until end */

tptr = tptr->link;

if (mark_and != and_head)

/* Undefined variables found */

terror (ERR_ce, ptr->ln);

/* Check all new funcn & data defs */

/** Check all new funcn & data defs */

/** Check all new funcn & data defs */

/** Check all new funcn & data defs */

/** Check all new funcn & data defs */

/** Check all new funcn & data defs */

/** Check all new funcn & data defs */

/** Check all new funcn & data defs */

/** Check all new funcn & data defs */

/** Check all new funcn & data defs */

/** Check all new funcn & data defs */

/** Check all new funcn & data defs */

/** Check all new funcn & data defs */

/** Check all new funcn & data defs */

/** Check all new funcn & data defs */

/** Check all new funcn & data defs */

/** Check all new funcn & data defs */

/** Check all new funcn & data defs */

/** Check all new funcn & data defs */

/** Check all new funcn & data defs */

/** Check all new funcn & data defs */

/** Check all new funcn & data defs */

/** Check all new funcn & data defs */

/** Check all new funcn & data defs */

/** Check all new funcn & data defs */

/** Check all new funcn & data defs */

/** Check all new funcn & data def...**/

PHITYPE
typetimes (ptr)

/** Semantic check "*" when used */

/** for types */

nodal ptr;
{extern void putform ();

PHITYPE type;

putform (semcheck (ptr->lptr));

/** Attach formal type to */

/** formal list */

/** Look for right type; if C, */

/** end of insertions */

putform (type);

/* Always return NULL; */

/* This value is used by pare.:

118
This file contains the following modules for the PHI parser:

* Matchfor
  * Tfunauxdef
  * Tfunid
* Tactualist
  * Tid
  * Telist

Algorithm:
This module contains the procedures needed to define and call functions. Tfunauxdef will set up the run-time structure of the function, Tfunid will check the semantics of the function, and matchfcr, called by tfunid, checks for the proper type & number of formal parameters.

Tactualist coordinates the checking of a function call. It uses both telist and act_walk. Act_walk determines whether the number & type of actuals is correct, and telist checks each element list and returns its type.

Tid performs semantic checking for program variables.

Externals

```c
#include <seanchec.h>
#include <string.h>

extern node types [];
extern varptr varhead;
extern void terror (), c_store_code ();
```

Globals

```c
int actual_count = 0; /* count of all actuals */
```

```
**** Matchfor **********

FLAG
  * matchfor (rptr, def)
    * total nptr;
    * defptr def;
  * extern long curr_addr;
  * extern node *getfptr ();
  * extern FLAG form;
  * node *nptr = getfptr (def);
    * form = FIND;
    * rptr = def->rptr;
    * curr_addr = 0;
```
if (nptr->name == IDENTIFIER) |
  (nptr->type) = tptr->type;
  nptr->addr = curr_addr;
  putvar (tptr->type, nptr);
  nptr = nptr->rptr;
  tptr = tptr->link;

else : /* Multiple forms */

  do
    nptr->lptr->type = tptr->type;
    nptr->lptr->addr = curr_addr;
    curr_addr = curr_addr +
     types !tptr->type...bytes;
    putvar (tptr->type, nptr->lptr);
    nptr = nptr->rptr;
    tptr = tptr->link;
  while ((nptr->=NULL) & (tptr->=NULL)); /* Halt when end reached */
  /* by either ptr */

  form = FALSE;

  if (nptr = NULL) tptr = NULL /* One ptr isn't at end of fur */
  return (FALSE);
  else return (TRUE);

/**************************** Ifunauxdef ******************************/

void
ifunauxdef (ptr)
  nodal ptr;
  /* Type check ifunauxdef */

  extern long curr_addr;
  extern void c_end proc (); c_end ();
  extern char *name ();
  extern nodal numconvert ();
  char *tmp = malloc (8);
  FLTYPE *type, rtype;
  varptr varl, mark = varhead;
  varp pres_addr = curr_addr;

  tmp = strcpy (name, name ());
  c_end (tmp);
  var = secheck (ptr->ptr);
  rtype = secheck (ptr->rptr);
  while (varhead->link = mark) ;
    varl = varhead;
    varhead = varhead->link;
    var->link = NULL;
    free (var);
    ptr->rptr =
    tr_numconvert (ltype, rtype, ptr->rptr);
  c_end proc (tmp);
  curr_addr = pres_addr;

  /* Reset addresses */
/* Semantic Check for tfunid */

馮 (ptr);
exect def* ptr. fnddef (ptr);
exect long struct addr;
exect str *name (ptr);
ct count = 0;
def ptr def;
char *type = "allen ("");

if (!def = fnddef (ptr -> ptr -> index))
    error (ERR_ff, ptr -> n);
    return (NOTFOUND);

else
    ptr = ptr -> type = def -> type;
    ptr = type = def -> type;
    putvar (ptr -> ptr -> type, ptr -> ptr, FALSE);
    if (!matchfor (ptr -> ptr -> def))
        error (ERR_fq, ptr -> n);
    else
        holder = strcopy (holder, name (ptr));

        while (* (holder -> count) + 1)
            * (ptr -> ptr -> label. count) =
            * (holder -> count) + 1;
            ++ count;
        ptr = ptr -> addr = 0;
        t_start_proc (ptr -> ptr -> label);

        return (ptr -> type);

/* Semantic Check for element list */

if (ptr -> ptr -> NULL)
    setcheck (ptr -> ptr);
setcheck (ptr -> 1.ptr);
store_code ("call ppop\n");
store_code ("push eax\n");
store_code ("push d.flf\n");
*actual_count;

/* Recursive procedure */

act .walk (ptr, fptr)
nodal ptr;
fnode *fptr;

if (ptr->rptr != NULL)
    act_walk (ptr->rptr, fptr->link);  /* Recurse until NULL ptr is hit */

semcheck (ptr->lptr);
if (ptr->lptr->name == ELIST)  
    ++actual_count;

c_store_code("call pop\n");
c_store_code("push cx\n");
c_store_code("push dx\n");

/*********************** Tactuals **********************/

PHITYPE

phit (ptr)  /* Evaluate actualists */
    nodal ptr;
    (extern void c_call_proc ();
    extern FLAG and_flag;
    extern varptr findvar ();
    extern defptr finddef ();
    extern char *name ();
    defptr def = finddef (ptr->lptr->index);  /* Defstack pointer */
    varptr var = findvar (ptr->lptr->index);  /* Varstack pointer */
    int count_hold = actual_count;  /* Buffer for long to string conv */
    char *long_buff = malloc (10);  /* Conversion variable */
    node *fptr;
    actual_count = 0;
    
if (def)  /* Definition found */
    
    if ((!var || !var) || var)  /* Legitimate cases */
        {  /* Get a ptr to the formal nodes */
            fptr = def->fptr;
            act_walk (ptr->rptr, fptr);
            convert = actual_count;
            c_store_code("mov bx, ");  /* Generate code to put # of */
            stcl_d (long_buff, convert);
            c_store_code (long_buff);
            c_store_code("\n");
            c_call_proc ("i_mov");
            
            if ((and_flag) || (!var))  /* Cover "and" scoping rules */
                add_and (ptr->lptr);
                c_call_proc (name ());  /* Holder for real name */
            else
                c_call_proc (var->nptr->label);  /* Gen code to call function */
                actual_count = count_hold;  /* Restore actual count */
                return (def->type);  /* Function name not found */
        }
        terror (ERR_nh, ptr->ln);
        return (NOTFOUND);
    
122
/**
 * PHITYPE
 */

tid (ptr) {
    nodal ptr;
    extern void s_u_form ();
    extern long d_v_r_add;
    extern char *name ();
    extern int formal ();
    extern char *flag;
    extern struct variable;
    extern struct def; /*
    char *long_puff = malloc ();
    struct var = malloc (ptr->type);
    before def;

    if (var)
        if (def = finddef (ptr->type)) {
            name (ptr);
            add (ptr);
            c_call_proc (name ());
            return (gettype (def));
        } else return (NOTFOUND);

    else if (formal (var))
        s_u_form (long_puff, var->v_r_add);
        c_u_form (long_puff);
    else
        c_call_proc (var->var->flag);
    return (settype (var));
PUBLIC DOMAIN SOFTWARE

Name: Semcheck Module #3
File: Sem3.c
Authors: Maj E.J. COLE / Capt J.E. CONNELL
Started: 01/02/87
Archived: 04/02/87

This file contains the following modules for the PHI parser:

Trdivide  Tidivide  Tarithop
Tprimary  Tconvert  Tconstant
Tand      Tor    Tnegation

Algorithm:
This module contains the procedures necessary for implementing
arithmetic & boolean operators. Tarithop coordinates the semantic
checking of arithmetic ops by calling the proper function based
on the operator type. Trdivide & Tidivide handle semantic checking
for real & int division, respectively. For all other arithmetic
ops, the numconvert procedure (sem0) is called to perform seman-
tic checking, then code is generated.
For each boolean operator, the appropriate child(ren) is checked
and code is generated for the operation.
In addition, tconstant checks the type of a simple constant by
calling convert, & then returns either the constant type or an error.

extern void terror ();
extern void c_store_code (); /* Store asm language output */

---

void trdivide (ptr);
/* Division of real operands */
range ptr;
PHITYPE *type, *rtype;
EXTERN FLAG err_found;
EXTERN void c_store_code (); /* Store asm language output */

switch (*type = semcheck (ptr->lptr)); /* Check left side for type */

switch (*type):
    case (REAL) : break;
    case (INTEGER):
    case (NATURAL):
        c_store ();
        break;
    default: terror (ERR_aa, ptr->., ptr->lptr); /* Lt child must ptr numeric type */
        return;
    /* Error, no need to go thru accessor */

124
rtype = semcheck (ptr->rptr);  // Check right side for type

switch (rtype) {
    case (REAL) : break;
    case (INTEGER) :
        c_f0r ();
        break;
    default : terror (ERR_aa, ptr->rptr->ln);
        return;  // Error, no need to go thru acode
}
acode (ptr, REAL);  // Generate code

/**************************************************
PHIYPE
endifdivr (ptr)  // Semcheck for integer division */
node ptr;
(PH1TYPE itype, rtype, type = NATURAL;
    itype = semcheck (ptr->lptr);
    rtype = semcheck (ptr->rptr);

switch (itype) {  // Check it for Int/Natural Type
    case (INTEGER) : type = INTEGER;
    case (NATURAL) : break;
    default : terror (ERR_cc, ptr->lptr->ln);
        return (INTEGER);
}

switch (rtype) {  // Check rt for Int/ Natural Type
    case (INTEGER) : type = INTEGER;
    case (NATURAL) : break;
    default : terror (ERR_cc, ptr->rptr->ln);
        return (INTEGER);
}
acode (ptr, type);  // Generate code

return (type);

/**************************************************
PHIYPE
arithop (ptr)  // Type Check Addition, Multiplication, Sequence Ops
node ptr;
(extern PHIYPE numconvert ());
int type;

switch (ptr->name) {
    case (ADD_) :  // Addition fails through
        aclode (ptr, type);
        return (type);
    case (SUB_) :  // Subtraction fails through
        aclode (ptr, type);
        return (type);
    case (MULT) :  // Mult. fails through
        aclode (ptr, type);
        return (type);
    else {
        terror (ERR_aa, ptr->ln);
        return (NATURAL);
    }

}
case (RDIV_) : tdivide (ptr);
    ptr->type = type;
    return (REAL);
  case (IDIV_ ) : tidivide (ptr);
    ptr->type = type;
    return (INTEGER);
  case (COLON_ ) : break;
    /* Dummies for now, */
    /* but watch our smoke!! */
    /* " " " " */
  case (CAT_ ) : break;
    
}                                                

/******************************************** Tprimary ********************************************/ 
PHITYPE tprimary (ptr) /* Handle unary "-" or ":" */
    nodal ptr;
{PHITYPE type;
    type = semcheck (ptr->rptr);
    if (type == INTEGER) &&
        (type == REAL) &&
        (type == NATURAL) /* Check type of right node */
        error (ERR_aa, ptr->rptr->ln); /* Type must be a number */
    else if (ptr->name == NEG_) { /* Negate operation */
        c_store_code("call igetvalue\n"); /* Spew code */
        c_store_code("neg ax\n");
        c_store_code("call iputvalue\n");
        return (type); /* Note that no action is req */
        /* for unary "-" */
    }

/******************************************** Convert ********************************************/ 
PHITYPE convert (string) /* Convert const to real, boolean, */
    stg string;
    {FLAG e = FALSE, /* String to convert */
      period = FALSE; /* True if "e" or "E" read */
      int count = 0; /* True if a period has been read */
      /* Garden variety loop counter */
      if (strcmpi (string, "FALSE")
          && strcmpi (string, "TRUE"))) { /* If not boolean */
          
          while (string [count] != 0) { /* Loop until end of string */
              if (!isdigit (string [count])) { /* If character is not a digit */
                  if ((string [count] == 'e') ||
                      (string [count] == 'E')) { /* "e" or "E" found */
                      if (e) return (ERROR); /* Cannot have two "e"s */
                      else {
                          e = TRUE;
                          if ((string [count + 1] == '+') ||
                              (string [count + 1] == '-')) ++count;
                      }
                      
                      }
                      
                      }
                      
                      }
                      
                      }
                      
                      }
```c
else if (string['count'] == '.' || isdigit(string['count'])) { /* Decimal point found */
    if (period) return (ERROR); /* Cannot have two periods */
    else period = TRUE;
}
else return (ERROR);

--count;

if (e period) return (REAL); /* If exponent has been read */
if (string['count'] == '-') return (INTEGER); /* Negative sign makes an integer */
return (NATURAL); /* If no other sum types, natural */
return (BOOLEAN); /* If not a number, a boolean */

/**************************** TConstant ***************************/

PHITYPE
tconstant (ptr)
    nodal ptr;
    extern put_addr ();
    PHITYPE type;
    NameRec *tptr;

    tptr = ptr->index;

    if (type = convert (tptr->name + 1)) { /* Calculate type */
        ptr->type = type;
        put_addr (ptr, type);
        c_l_const (tptr->name + 1);
        return (type);
    } /* No legitimate constant found */

/**************************** Tand ***************************/

PHITYPE

tand (ptr)
    nodal ptr;
    PHITYPE ltype, rtype;

    ltype = semcheck (ptr->lptr);
    rtype = semcheck (ptr->rptr);

    if ((ltype == BOOLEAN && rtype == BOOLEAN)) { /* Both children must be boolean */
        terror (ERR_kk, ptr->ln);
        c_store_code ("call land\n");
        return (BOOLEAN);
    }

/**************************** Tor ***************************/

PHITYPE
tor (ptr)
    nodal ptr;
    PHITYPE ltype, rtype;

    ltype = semcheck (ptr->lptr);
    rtype = semcheck (ptr->rptr);

    if ((ltype == BOOLEAN && rtype == BOOLEAN)) { /* Both children must be boolean */
        terror (ERR_kk, ptr->ln);
        c_store_code ("call lor\n");
        return (BOOLEAN);
    }
```

127
c_store_code("call ior\n");
return (BOOLEAN);

/**************************** Tnegation **************************/

PHETYPE

tnegation (ptr
  local ptr;
  if (!isemcheck (ptr->rptr == BOOLEAN))
    error (ERR_X, ptr->ln);
  else c_store_code("call nior\n");
  return (BOOLEAN);
PUBLIC DOMAIN SOFTWARE

Name: Semcheck Module #4
File: Sem4.c
Authors: Maj E.J. COLE / Capt J.E. CONNELL
Started: 01/29/87
Archived: 04/03/87
Modified:

This file contains the following modules for the PHI compiler:

Tif
Then
Telseif
Telse
Tcomp

Algorithm:
This module contains the procedures necessary to implement the "if-then-elseif-else" series of commands. Tif coordinates the semantic checking by calling Then to check its left nodes, then calling Telse to check its right nodes. Telse will be called until the right subtree runs out of "elses" and "elseifs".

Modified:

extern FLAG err_found;
extern PHIYPE semcheck ();
extern PHIYPE numconvert ();
extern char *name ();
extern void terror (), c_store_char ();

char *if_label = NULL;

PHITYPE
Tif (ptr)
{
    nodal ptr;
    extern PHIYPE numconvert ();
    PHIYPE type;
    if (if_label == NULL) if_label = malloc (8);
    if_label = strcpy (if_label, name ());
    type = numconvert (ptr);
    c_store_code (if_label);
    c_store_code (":\n")
    return (type);
}
/********************** Tthen ***********************/

PHITYPE
tthen (ptr)

    nodal ptr;
    (PHITYPE ltype, rtype);
    char *label = calloc (7,1);
    char *holder = calloc (7,1);

    strcpy (holder, if_label);
    if((ltype = semcheck (ptr->lptr) ) ! BOOLEAN) */ Left node contains condition; */
        terror (ERR_11, ptr->lptr->ln);
        if_label = strcpy (if_label, holder);
    label = strcat (label, name ());
    c_store_code ("call igetvalue\n");
    c_store_code ("cmp ax,1\n");
    c_store_code ("jne ");
    c_store_code (label);
    c_store_code ("\n");

    rtype = semcheck (ptr->rptr); */ Check right side */
    c_store_code ("jmp ");
    c_store_code (if_label);
    c_store_code ("\n");
    c_store_code (label);
    c_store_code (":\n");

    return (rtype); */ Right type is returned */

/******************** Telseif **********************/

PHITYPE
telseif (ptr)

    nodal ptr;
    (PHITYPE numconvert ());

    return (numconvert (ptr));

/******************** Telse *********************/

PHITYPE
telse (ptr)

    nodal ptr;

    return (semcheck (ptr->lptr)); */ Return left side; */
    /* right side is always else */

/******************** Tcomp **********************/

PHITYPE
tcomp (ptr)

    nodal ptr;
    (PHITYPE numconvert ());
    PHITYPE type;

130
type = numconvert (ptr);

switch (ptr->name) {
    /* Check cases */
    /* WORKS ONLY FOR INTEGERS AND */
    /*BOOLEANS --- NEEDS REAL */
    case (EQ) : c_store_code ("call leq\n");
            break;
    case (NEQ) : c_store_code ("call ineq\n");
            break;
    case (KW_\_LESS) :
        c_store_code ("call ilt\n");
            break;
    case (KW_\_GREATER) :
        c_store_code ("call iqt\n");
            break;
    case (LEQ) : c_store_code ("call ilteq\n");
            break;
    case (GEQ) : c_store_code ("call igt\n");
            break;
    case (KW_\_IN) :
        c_store_code ("call in\n");
            break;
    case (KW_\_NOTIN) :
        c_store_code ("callnotin\n");
            break;
    default :
        terror (ERR_\_1, ptr->ln);
            break;
    }
    return (BOOLEAN);
}
PUBLIC DOMAIN SOFTWARE

Name: Semcheck Utilities
File: Sem_U.c
Authors: Maj E.J. COLE / Capt J.E. CONNELL
Started: 01/02/87
Archived: 04/03/87
Modified:

This file contains the following modules for the PHI parser:

- Putvar
- Putform
- Makeform
- Findvar
- Getfptr
- Getvtype
- Finddef
- Put_addr
- Name
- Getdtype
- Form
- Makevar
- Putdef
- And_alloc
- Add_And
- Del_And

Modified:

******************************************** Externals ********************************************
#include <semcheck.h>
#include <string.h>
/* for "stpcpy"

******************************************** Globals ********************************************
FLAG err_found = FALSE;
/* True if an error found
long curr_addr = START_ADDR;
/* Next address to be used to place a variable
long curr_scope = START_ADDR;
/* Current scope
form = FALSE;
/* True if formals being processed

******************************************** Typetable Definitions ********************************************
int typeptr = TYPE_INIT;
/* Ptr to last typetable insert
node types [MAXTYPES];
/* Typetable

******************************************** Vartable Definitions ********************************************
varptr varhead = NULL;
/* Head of varlist linked list

******************************************** Deftable Definitions ********************************************
defptr defhead = NULL;
/* Head of deftable linked list

******************************************** And_List Definitions ********************************************
and_ptr and_head = NULL;
/* Head for and list
and_flag = FALSE;

******************************************** Makeform ********************************************

fnode

**makeform ()

/ * Create a formal node

  return ((fnode*) calloc (1, sizeof (fnode)));

******************************************** Putform ********************************************

void
putform (type)

/* Put type into formal list
    PHITYPE type;
    extern fnode *head;
    fnode *ptr = **makeform () ;
    **tracer;
/* Make a formal node
/* Tracer for the formal list

132
ptr->type = type;

if (ptr->head != NULL) {
    tracer = ptr->head;
    while (tracer->link != NULL)
        tracer = tracer->link;
    tracer->link = ptr;
    ptr->link = NULL;
} else {
    ptr->head = ptr;
    ptr->link = NULL;
}

/*********************** Makevar ***********************/
varptr
makevar ()
    return (struct varnode*)
        malloc (1, sizeof (struct varnode));

/*********************** Putvar ***********************/
void
putvar (type, treenode)

    return;

    ptr->nptr = treeode;
    ptr->type = type;
    ptr->form = form;
    ptr->link = varhead;
    varhead = ptr;
    ptr = NULL;
    free (ptr);

/*********************** Findvar ***********************/
varptr
findvar (varname)

    while (ptr != varhead) {
        if (ptr->nptr->index == varname)
            return (ptr);
        ptr = ptr->link;
    }
    return (NULL);
/* Getvtype */

PHITYPE
getvtype (ptr)

varptr ptr;
{
    return (ptr->type);
}

Putdef

void putdef (type, treeptr)

PHITYPE type;
modal treeptr;
extern int form;

defptr ptr = (struct defnode*)malloc(sizeof (struct defnode));

ptr->mode = treeptr;
ptr->type = type;

ptr->link = defhead;
defhead = ptr;
ptr = NULL;
free (ptr);

Finddef

defptr finddef (varname)

long varname;
defptr ptr = defhead;

while (ptr != NULL) {
    if (ptr->mode->index == varname)
        return (ptr);
    ptr = ptr->link;
}

return (NULL);

getfptr

inode "getfptr (ptr)

defptr ptr;

return (ptr->fptr);

Getdtype

PHITYPE
getdtype (ptr)
defptr ptr;

return (ptr->type);

Add_and

void add_and (ptr)
modal ptr;

/* Add and_node to and list */

/* Ptr to node containing var */
extern and_ptr and_head, and_alloc ();
extern int buff_ptr;
and_ptr a_ptr = and_alloc (); /* Holder for and_ptr */
a_ptr->buff_ptr = buff_ptr; /* Set ptr to current buffer ptr */
a_ptr->ptr = ptr; /* Set ptr to node with var set */
a_ptr->link = and_head; /* Link node to list */
and_head = a_ptr; /* Dispose of a_ptr */

a_ptr = NULL;
free (a_ptr);

//***************************** And_Alloc *******************************/
and_ptr
and_alloc () /* Create a node for and_list */
return ((struct and_struct*)malloc (1, sizeof (struct and_struct)));

//***************************** Del_and *******************************/
void
del_and (ptr) /* Delete entry into the and_list */
and_ptr ptr;
:extern and_ptr and_head;
and_ptr search = and_head;

if (ptr != and_head) i
/* Case if pointer not equal to */
/* first entry in list */

while (search->link != ptr) /* Place ptr on entry above */
/* Get entry */

search = search->link;

search->link = ptr->link; /* Set pointer */

else and_head = ptr->link; /* Case ptr = to list entry in list */

ptr->link = NULL; /* Dispose of unused node */
free (ptr);

//***************************** Terror *******************************/
void
terror (err_num, line_num) /* Sem check error handling */
/* routine */
int err_num, line_num;
:extern ErrorHandler ();

err_found = TRUE; /* Set err_found to true */
/* stop code gen */
/* generic error handling stmt */

ErrorHandler (line_num, err_num, SEM_ERR);

//***************************** Putaddr *******************************/
void
put_addr (ptr, type) /* Inserts virtual address */
/* variable function return */
/* And increments cur_addr */
/* Assumes normal function */
/* Pointer to target */

nodal_ptr;
{
  PHITYPE type;

  if (curr_addr > MAXADDR)
    terror (ERR_mm, ptr->ln);

  /* Node type */
  ptr->addr = curr_addr;
  ptr->scope = curr_scope;
  curr_addr = curr_addr + (types[type].bytes);
  /* Increment curr_addr by num of bytes type needs */

  /* Set node address */

  if (curr_addr > MAXADDR)
    /* Error if address exceeds address space */
    terror (ERR_mm, ptr->ln);

  /* Generate an appropriate name for a label/procedure */
  char *name ()

  char *string = malloc (7),

  /* Holder for output */
  *string1 = malloc (7);

  /* Number to append to string */
  static long seed = 10000;

  /* String prefix */
  *string = 'a';
  *(string - 1) = ENDSRING;
  stcd (string1, seed);
  string = strcat (string, string1);

  /* Insert string terminator */
  --seed;

  /* Convert long seed to string */
  return (string);

  /* Concatenate strings */

  /* Incr int to avoid duplication */

  /* Returns true if the varnode describes a formal */
  FLAG formal (ptr)

  varptr ptr;

  if (ptr->form) return (TRUE);

  /* Returns true if the varnode */
  else return (FALSE);

  /* describes a formal */
APPENDIX K
ROCK COMPILER — CODE GENERATION MODULE

* PUBLIC DOMAIN SOFTWARE
* Name : Code Generation Module
* File : Code_Gen.c
* Authors : Maj E.J. COLE / Capt J.E. CONNELL
* Started : 02/06/87
* Archived : 04/10/87
* Modified : 04/13/87 Code output to vdisk

This file contains the following modules for the PHI compiler:

- **C_Store_Code**
- **C_Startup**
- **C_Start_Up**
- **C_Printcode**
- **C_Jmp**
- **C_Start_Proc**
- **C_I_Const**
- **C_I_Form**
- **C_End_Proc**
- **C_Call_Prog**

Algorithm:
This module contains the procedures necessary for code generation.

- **C_Startup** initializes the run-time file, & the semantic checker, & then call the procedures as necessary. Note that "C_store_code" is a generic generator which will spew any string given as an arg to the output file.

Modified : 04/13/87 Code output to vdisk, drive "a:"
if (err_found) ; /* Compute only if no error found */
while (string + ptr) != NULL) ; /* Copy string char by char */
  code_buffer + buff_ptr = *(string + ptr);
  ++ptr;
  ++buff_ptr;

/********************************************** C_Jmp ***********************************************/
void
c_jmp (name) ; /* Gen code to insert jump command */
  char *name;
  c_store_code ("jmp ");
  c_store_code (name);
  c_store_code ("\n");

/********************************************** C_Start_Proc ***********************************************/
void
c_start_proc (name) ; /* Output name for start of asm */
  char *name;
  c_store_code (name);
  c_store_code ("\n");

/********************************************** C_End_Proc ***********************************************/
void
c_end_proc (name) ; /* Output name for ending an */
  char *name;
  c_store_code ("call del_scope\n");
  c_store_code ("ret\n");
  c_store_code (name);
  c_store_code ("\n");

/********************************************** C_Call_Proc ***********************************************/
void
c_call_proc (name) ; /* Output call for an assembly */
  char *name;
  c_store_code ("call ");
  c_store_code (name);
  c_store_code ("\n");

/********************************************** C_I_Form ***********************************************/
; /* Generate call to put integer */
  ; /* format add onto stack */
  c_store_code ("mov r2,\n");
  c_store_code (name);
  c_store_code ("\n");
  c_store_code ("call \form, ");
/*******************************************************************************/
void
c_i_const (name) /* Output code for assigning an integer constant */
{
    char *name;
    c_store_code ("mov ax,"");
    c_store_code (name);
    c_store_code ("n");
    c_store_code ("call i4putvalue\n");
}
/*******************************************************************************/
void
c_i_op (op) /* Output code for int arith ops */
{ extern void terror();
    switch (op) {
        case (ADD) : c_call_proc ("iadd");
                    break;
        case (SUB) : c_call_proc ("isub");
                    break;
        case (DIVIDE) : c_call_proc ("idivn");
                    break;
        case (MULT) : c_call_proc ("imult");
                    break;
        default : return;
    }
}
/*******************************************************************************/
void
c_startup () /* Open and initialize files */
{ code_buffer = getmem (SIZEBUFFER);
    c_store_code ("extrn initial : near\n");
    c_store_code ("extrn ladd : near\n");
    c_store_code ("extrn isub : near\n");
    c_store_code ("extrn imult : near\n");
    c_store_code ("extrn idivn : near\n");
    c_store_code ("extrn lequ : near\n");
    c_store_code ("extrn lineq : near\n");
    c_store_code ("extrn lgt : near\n");
    c_store_code ("extrn llt : near\n");
    c_store_code ("extrn land : near\n");
    c_store_code ("extrn lor : near\n");
    c_store_code ("extrn igteq : near\n");
    c_store_code ("extrn iptime : near\n");
    c_store_code ("extrn ilteq : near\n");
    c_store_code ("extrn igvalue : near\n");
    c_store_code ("extrn init : near\n");
    c_store_code ("extrn finis : near\n");
    c_store_code ("extrn print_top : near\n");
    c_store_code ("extrn negation : near\n");
    c_store_code ("extrn iformal : near\n");
    c_store_code ("extrn imov : near\n");
    c_store_code ("extrn ppush : near\n");
    c_store_code ("extrn pop : near\n");
    c_store_code ("extrn add_scope : near\n");
    c_store_code ("extrn del_scope : near\n");
    c_store_code ("org 0100h\n");
    c_store_code ("cseq\n");
    c_store_code ("call initial\n");
}
/*************** C_Print_Code ***************
void c_print_code ()
{ extern char prefix[]; int code;
char holder[30];
strcpy (holder, "d:");
strcat (holder, prefix);
strcat (prefix, holder);
strcat (holder, "a.86");

code = open(FILENAME,O_TRUNC | O_WRONLY,NULL); /* Open file for writing and */
write (code, code_buffer, buff_ptr); /* Write the buffer */
close (code); /* Close the output file */
}

/*************** C_Ending ***************
void c_ending ()
{ if (!err_found) {
c_store_code ("call print_top\n"); /* Print address pointed to by */
c_store_code ("call finis\n"); /* Routine to make clean ending */
*(code_buffer + (buff_ptr ++)) = CNTRLZ; /* If no error, put asm language */
c_print_code (); /* delimeter to file */
c_store_code (); /* Output code to a file */
}

/*************** C_ztor ***************
void c_ztor ()
{ extern void terror ();
int name;

name = ptr->name;

switch (name) {
case (ADD_) : if (type == REAL); /* Addition */
else c_i_op (ADD); break;
case (SUB_) : if (type == REAL); /* Subtraction */
else c_i_op (SUB); break;
case (MULT_) : if (type == REAL); /* Multiplication */
else c_i_op (MULT); break;
case (RDIV_): break;
}

*************** Acode ***************
void acode (ptr, type)
{ extern void terror ();
int name;

name = ptr->name;

switch (name) {
case (ADD_) : if (type == REAL); /* Addition */
else c_i_op (ADD); break;
case (SUB_) : if (type == REAL); /* Subtraction */
else c_i_op (SUB); break;
case (MULT_) : if (type == REAL); /* Multiplication */
else c_i_op (MULT); break;
case (RDIV_): break;
}
case (IDIV) : c_i_op (DIVIDE);
    break;
    /* Integer Division */
APPENDIX L

ROCK COMPILER — USER INTERFACE

//***************************************************************
* Name : User Interface
* File : User.C
* Authors : Maj E.J. COLE / Capt J.E. CONNELL
* Started : 04/01/87
* Archived : 04/10/87
* Modified :
***************************************************************

* This file contains the following modules for the PHI compiler

* User_err Getname Prog_name
* Print_header P_Close User

* Algorithm :
* This module contains the procedures necessary for the user inter-
* face.
* Prog_Name gets the user's choice of program by calling Get_Name
* Print header is called to print the initial screen display on con-
* sole, & the User procedure is the overall coordinator of the inter-
* face.
* User_Err and P_Close are both independent procedures. User_Err
* handles output in the event that an error or errors have been found.
* P_close is called by "Rock_Main" to ensure the input file has been
* closed.

***************************************************************
* Modified :
***************************************************************

/*************************************************************** Externals ***************************************************************/
#include <user.h>
#include <dos.h>
#include <stdio.h>
extern void clrscr (), mov_cursor (), cir_window ();

/*************************************************************** Globals ***************************************************************/
char s_name [BUFFLENGTH],
    prefix [BUFFLENGTH];
FILE *infile;

/*************************************************************** User_Err ***************************************************************/
void
user_err ();
extern void clrscr ();

142
extern int num_errors;
FILE *errors;
int numblocks,
count = 0;
char *buffer = malloc (BSIZE),
input;

errors = fopen (ERRORFILE, "a");
printf("Number of errors = %d\n",num_errors);
puts("$", errors);
close (errors);

cirscnt();
errors = fopen (ERRORFILE, "r");

numblocks = fread (buffer,BLOCKSIZE,20,errors); /* Read error msgs from error files*/
/* BLOCKSIZE will allow whole */
/* file to be read at once */

while (*(buffer + count) != '$') {
  putchar (*(buffer + count));
  --count;
}

printf("\n \n \n\n");
printf("%s", PAUSE);

input = getch ();
close (errors);
cirscnt();

if (input == ESCAPE) exit (1);

/*******************************************************************************/

Getname

Returns the user’s choice of file to compile
Single input character Buffer pointer

/* Loop, get file name itr by itr */
/* <- key is hit */
/* Backspace */
/* Insert blank */
/* Eat last char if there is one */

if (input == ESCAPE) {
  cirscnt ();
  exit (1); }
else if (ch < 127) {
  putchar (ch);
  name [count] = ch;
  ++count; }
while ((count <= BUFLENG74) &&
  ch != EOLN);
void prog_name ()
{
  do :
    clr_window (9,1,21,79);
    mov_cursor (10,2);
    printf (GETPROGRAM);
    getname ();
    infile = fopen (u_name, "r");
    if (! infile)
    { /* Name not in current directory */
      mov_cursor (20,33);
      printf (FILE1_ERROR);
      mov_cursor (21,16);
      printf (FILE2_ERROR);
      if (getch () == ESCAPE) 
        clrscr (); 
        exit (1); 
    }
    while (! infile);
    mov_cursor (13,28);
    printf (WAIT);
}

void print_header ()
{
  clrscr ();
  mov_cursor (1,33);
  printf (HEADER1);
  mov_cursor (2,24);
  printf (HEADER2);
}

void p_close ()
{
  close (infile);
}

void user ()
{
  int count = 0;
  print_header ();
  prog_name ();

  while (strlen (u_name [count]) != NULL) 
    
  prefix [count] = u_name [count] ;
  *count ;
  prefix [count] = 0 ;
  
  /* Insert end of string char */

  /* Get legitimate program name */
  /* Loop until fopen finds a legit name */
  /* Clear out lower window of screen */
  /* Name not in current directory */
  /* Print user friendly error msgs */
  /* Exit if ESCAPE pressed */
  /* Repeat until correct file found */
  /* NOTE - escape exits loop & prints */
  /* Print out header for user */
  /* Close out target file */
  /* Invoke user interface */
  /* Duty integer */
  /* Copy root of input file name */
  /* Loop until end of input name */
  /* reached OR until end of string */
  /* reached, if no extension */
  /* Insert end of string value */

144
APPENDIX M
ROCK COMPILER — RUNTIME UTILITIES

;***********************************************************************
;* Name : Phi Runtime Utilities
;* File : U.a86
;* Authors : Maj E.J. COLE / Capt J.E. CONNELL
;* Started : 01/26/87
;* Archived : 16 Feb 87
;* Modified : 16 Apr 87 Stack/VarSpace Crash error check EC
;***********************************************************************

;***********************************************************************
;ALGORITHMS
;***********************************************************************

1. Input/Output: The first section of the program contains input and output.

2. Virtual Space: A virtual space is set up in the extra segment to hold both the stack. The middle of this space is denoted by the symbol "vars", and variables offset (+32700) from vars. In this implementation, the program stack grows from vars grow from the bottom. The virtual space is assumed to be made up of words (two bytes), so only even numbers may be used to access it.

3. Stack: The stack pointer is the si register, which is initialized to 32700. As the stack grows, the si register is reduced by two. Push and pop will push and pop two registers. "Push_one" and "Pop_one" will push and pop single words to and from the stack.

4. Addressing Program Variables: Each program variable is assigned a two-tuple A scope and 0 is the offset from the base address of variables in that scope. The address of a variable given A.

5. Scoping: Initially the scope is set to 0: the global scope. The variable space containing the outer scope, and the variable "S_Nest" contains the current new scope is created, "S_Nest" is increased by one, and the three-tuple S = (L = Static Link, pointing nesting level of the outer scope, N is the nesting is the base address of display of variables for this scope. When a scope is deleted, the top of the stack is saved, the top instantiation of S and S_Link and S_Nest are recalculated.

6. Inserting/Extracting Program Variables: "I_Assign" will insert an integer or scope contained in S_Nest when it is requested. "IptutValue" will insert the corresponding tuple A on the stack. "Getvalue" will pop the tuple A off the top of the value of the integer pointed to by A.

;***********************************************************************
;***********************************************************************

* Modified : 22 Feb 87 Add/del_scope changed to save TOS. EC
* 16 Apr 87 Added check for stack/varspace crash, includes message to observer

145
Public Procedures

Public i_mov
Public i_formal
Public i_getvalue
Public finish
Public inputvalue
Public find_addr
Public add_scope
Public del_scope
Public initial
Public finis
Public ppush
Public pop
Public lassign
Public lor
Public land
Public leq
Public ineq
Public lit
Public lgt
Public lteq
Public lgteq
Public negation
Public ladd
Public isub
Public imult
Public idivn
Public print_top

I/O Procedures

Public _char

Print a char to the screen
assumes letter to be printed is in dl register

Print a char to the screen
assumes letter to be printed is in dl register

Print Num

Prints, as a number, the value found in the bx register
pr-:ntn~u':
push ax
push bx
push cx
push dx
mov cx, 10000
cmp dx, 0
jge small
mov dx, '-'
call print_char
neg dx

small:
 cmp bx, 10
 j1 final

div_loop: mov ax, bx
 xor dx, dx
 div cx
 cmp ax, 0
 jne p_loop
 mov ax, cx
 mov cx, 10
 xor dx, dx
 div cx
 mov cx, ax
 jmp div_loop

p_loop: mov ax, bx
 xor dx, dx
 div cx
 mov bx, dx
 add ax, 48
 mov dx, ax
 call print_char
 xor dx, dx
 mov ax, cx
 mov cx, 10
 div cx
 mov cx, ax
 cmp ax, 1
 jne p_loop

final:
 add bx, 48
 mov dx, bx
 call print_char
 call echo
 pop dx
 pop cx
 pop bx
 pop ax
 ret

;********************************************************** Print_top **********************************************************
;
;Prints the space pointed to by the top tuple of the program stack
;
;print_top: mov dl, si
 add dl, 1
 mov bx, var_s dl
 add dl, 2
 mov bx, var_s dl
 mov offset to bx
call find_addr
mov di, cx
mov bx, vars [di]
call print_num
call eol
ret

;******************************* print_s *******************************
;assumes address of is in the dx register
;assumes string ends with a "$" sign
;
print_s:
push ax ;save register
mov ah, 9
int 21h
pop ax
ret

;***********************************************
;Stack Procedures
;***********************************************

;******************************* Ppush ********************
;Pushes values from cx (offset) and di (nesting level)
;
Ppush:  mov vars [si], cx
        sub si, 2
        mov vars [si], di
        sub si, 2
        cmp si, curr_addr
        jg p_return
        mov dx, offset crash
        call print_s
        call finish
        p_return:  ret

;******************************* Push_one ********************
;Push a single integer from cx register to the program stack
;
Push_one:  mov vars [si], cx
           sub si, 2
           ret

;******************************* PPop ********************
;Pop values from the program stack to di (nesting level) and cx (offset)
;
PPop: add si, 2
      mov di, vars [si]
      add si, 2
      mov cx, vars [si]
      ret

;******************************* Pop_one ********************
;Pop a single integer from the stack to the cx register
;
Pop_one: add si, 2
         mov cx, vars [si]
         ret
Varspace Management Procedures

********** IAssign **********

;Assign an integer value to a variable space in current scope
;Assumes value is in ax; offset is set to current max offset

Iassign:

; get static link
sub dl, 2
mov dl, s_link
mov dl, var 'dl
add dl, max_offset
mov vars dl, ax
add max_offset, 2
add curr_addr, 2
ret

**************************************************************

************ IGetValue ************

;Pop the stack and move the integer value pointed to into the ax register

Igetvalue:
call ppop;
mov dx, dl
call find_addr
mov dl, cx
mov ax, var 'dl
ret

**************************************************************

************ Iputvalue ************

;Takes an integer from AX register, puts its value into varspace, then puts its address on the top of the stack

Iputvalue:

mov dx, s_nest
mov cx, max_offset
call find_addr
mov dl, cx
mov vars dl, ax
mov dl, s_nest
mov cx, max_offset
call ppush
add max_offset, 2
add curr_addr, 2
ret

**************************************************************

** Scoping Procedures **

**************************************************************

************ FindAddr ************

;Returns address of variable at nesting level dx, offset cx to cx reg

Find_addr:

mov dl, s_link
find_loop:
cmp es:vars dl, dx
le f_out
add di, 2
mov di, es:vars[di]
jmp find_loop

f_out: sub di, 2
add cx, es:vars[di]
ret

;******************* Add_Scope ***********************
;Start new scope by adding static link, starting address, & nesting level
;
add_scope: mov cx, s_link
inc s_nest
mov di, s_nest
call ppush
mov cx, curr_addr
call ppush
mov di, max_offset
call ppush
mov :ax offset, 0
mov s_link, si
add s_link, 6
ret

;******************* Del_Scope ***********************
;Deletes a scope
;
del_scope: call ppop;
mov cx, di
call find_addr
push cx
dec s_nest
mov si, s_link
sub si, 4
mov cx, es:vars [si]
mov max_offset, cx
mov bx, 2
mov cx, es:vars [si-bx]
mov curr_addr, cx
add si, 6
mov cx, es:vars [si]
mov s_link, cx
pop di
mov ax, es:vars [di]
call iputvalue
ret

;****************************************************************

;Begin/End Procedures

;****************************************************************

;*********************** Initial ***********************
;initialize the stack and variables
;must initialize cx to base of stack heap before calling this
;
initial: mov si, SPACE_TOP
mov di, 0
mov cx, 0
call ppush
ret

150
;***************************************************************************
finis***************************************************************************

finis:
im 21h
ret

;***************************************************************************
Booleans
;***************************************************************************

;***************************************************************************
Negation
***************************************************************************
Negates a boolean value

negation:
call igetvalue
    cmp ax, 1
    jne zero
mov ax, 0
jmp end
zero:
    mov ax, 1
p: call iputvalue
    ret

;***************************************************************************
Lor
***************************************************************************
Takes logical or of two booleans and stacks address of answer

lor:
call igetvalue
    cmp ax, 0
    jne req
mov bx, ax
    call igetvalue
    or ax, bx
    call iputvalue
    ret

;***************************************************************************
Land
***************************************************************************
Takes logical and of two booleans and stacks address of answer

land:
call igetvalue
    cmp ax, 0
    jne req
mov bx, ax
    call igetvalue
    and ax, bx
    call iputvalue
    ret

;***************************************************************************
Ieq
***************************************************************************
Takes logical equal of two integers and stacks address of answer

ieq:
call igetvalue
    cmp ax, 0
    jne req
mov bx, ax
    call igetvalue
    eq
    mov ax, bx
    call igetvalue
    ret

***************************************************************************
;******************************************************************************* Ineq *******************************************************************************

 Takes logical not equal of two integers and stacks address of answer

 .ineq: call .getvalue
        mov cx, ax
        call .getvalue
        cmp ax, cx
        je .ineq
        mov ax, FALSE
        ret

 .ineq: mov ax, TRUE
        ret

;******************************************************************************* Ilt *******************************************************************************

 Takes logical less than of two integers and stacks address of answer

 Returns true if first value is less than the second value

 .ilt: call .getvalue
        mov cx, ax
        call .getvalue
        cmp ax, cx
        jge .less
        mov ax, TRUE
        ret

 .less: mov ax, FALSE
        ret

;******************************************************************************* Igt *******************************************************************************

 Takes logical greater than of two integers and stacks address of answer

 Returns true if first value is greater than the second value

 .igt: call .getvalue
        mov cx, ax
        call .getvalue
        cmp ax, cx
        jle .greater than
        mov ax, TRUE
        ret

 .greater than: mov ax, FALSE
        ret

;******************************************************************************* Ilteq *******************************************************************************

 Takes logical \le of two integers and stacks address of answer

 Returns true if first value is less than or equal to the second value

 .ilteq: call .getvalue
        mov cx, ax
        call .getvalue
        cmp ax, cx
        jne .ilteq
        mov ax, TRUE
        ret

152
**Igteq**

Takes logical 2 of two integers and stacks address of answer

Returns true if first value is greater than or equal to the second value

```
;Igteq: call igetvalue
  mov bx, ax
  call igetvalue
  cmp ax, bx
  jl gteq
  mov ax, TRUE
  jmp con3
reteq: mov ax, FALSE
  jmp con3
reteq:
```

**Integer Operations**

**Iadd**

Adds two integer values

Assumes offset off second value is in SI register

Offset of first value is at the top of the stack

```
;iadd: call igetvalue
  mov bx, ax
  call igetvalue
  add ax, bx
  jf err
  call igetvalue
  ret
err:    mov ax, offset add_err
  call print_s
  call ecn
  call exit
reteq:
```

**ISub**

Subs two integer values

Assumes offset off second value is in SI register

Offset of first value is at the top of the stack

```
;isub: call igetvalue
  mov bx, ax
  call igetvalue
  sub ax, bx
  jf err
  call igetvalue
  ret
err:    mov ax, offset sub_err
  call print_s
  call ecn
  call exit
reteq:
```
jo errs
; if overflow, run time error
call inputvalue
ret
;
errs: mov dx, offset sub_err
; Print error message on overflow
call print_s
call eoln
call finish
ret

;************************************************ IMult ************************************************
; Multiplies two integer values
; Assumes offset off second value is in SI register
; Offset of first value is at the top of the stack
; imult:
call inputvalue
mov bx, ax
; First value to bx register
call inputvalue
imul bx ; Perform mul, result in AX
jc err1 ; if carry set, run time error
call inputvalue
ret
;
err1: mov dx, offset mul_err
; Put integer into varspace
call print_s
call eoln
call finish
ret

;************************************************ IDivn ************************************************
; Divides two integer values, result in varspace, address of result stacked
; Offset of first value is at the top of the stack
; idivn: push cx ; Save Registers
push dx
; Get divisor
call inputvalue
mov bx, ax ; Move divisor to bx
; Get dividend to ax
;
xor dx, dx
; Set dx to 0
mov cl, 1
mov ch, 1
cmp bx, 0 ; cl and ch are negative flags
; bx is positive, no action needed
je err2 ; bx is 0, ERROR
jne err2 ; bx is negative, cl flag negated
neg bx

err2: cmp ax, 0 ; test dividend
; dividend >= 0, no action
je dloop ; ax is negative, cl flag negated
neg ax

; Loop and count subtractions
sub ax, bx
cmp ax, 0
jl done
inc dx
jmp dloop

; Store result in dx
; continue loop

154
Function Calling Procedures

** Function Calling Procedures **

;i_mov ********** i_mov **********
i_mov:    pop ret_addr
         ;Save i_mov's return address
         call add_scope
         ;mov addresses to cx and dx: regs
         pop cx
         ;mov dx, es:vars [di]
         mov di, cx
         call find_addr
         ;Set virtual address of the integer
         mov dx, es:vars [di]
         call lassign
         dec bx
         cmp bx,0
         jne strt
         push ret_addr
         ;Restore i_mov's return address
         ret

;i_formal ********** I_formal **********
i_formal:  mov di,0
           ;mov di, s_rest[di]
           call ppush
           ;Push offset and rest into stack
           ret

Variables

;seq

155
**** Constants ****************************

TRUE EQU 1
FALSE EQU 0
SPACE_TOP EQU 32700 ; Top of memory space

**** Integer Variables ****************************

max_offset dw 0 ; Maximum current offset w/in scope
curr_addr dw -32700 ; Current maximum address
s_link dw SPACE_TOP ; Current address of static link
s_nest dw 0 ; Current static nesting level
ret_addr dw 0

**** Error Messages ****************************

div_err db 'DIVISION BY ZERO, FOOL!' db 'S'
mul_err db 'MULTIPLICATION OVERFLOW, IDIOT!' db 'S'
add_err db 'ADDITION OVERFLOW, DIMWIT!' db 'S'
sub_err db 'SUBTRACTION OVERFLOW, NITWIT!' db 'S'
crash db 'STACK/VARIABLE SPACE CRASH' db 'S'

e seq
vars dw 0
end

156
APPENDIX N – TEST SUITE

SIMPLE TESTS OF FUNCTIONS AND VARIABLES

let c : $Z -> $Z;
c (20) where c (n) == if 1 = 2 then 3 * n 
else 3 + n endif

-- Simple "Hello I'm Alive Test"

let c : $Z -> $Z;
c (1 * 2) where c (n) == n * 3

-- Test for expression in functions's formals

let c : $Z -> $Z;
c (k + 2) where k == 2 and 
c (n) == if n = 1 then n * 3 else n + 4 endif

-- Test for expression in function’s formals

TESTS FOR RECURSION

let c : $Z -> $Z;
c (k * 2) where k == 2 and c (n) == n * 3

-- Test for expression in function's formals

let c : $Z -> $Z;
c (0) where c (n) == if n = 0 then 1 else c (n - 1) * n endif

-- Test for recursion in functions

let c : $Z -> $Z;
c (5) where c (n) == if n = 0 then 1 else c (n - 1) * n endif

-- Test for recursion in functions

let c : $Z -> $Z;
c (3) where c (n) == if n = 0 then 1 else n * c (n - 1) endif

-- Test for recursion in functions

let c : $Z ->$Z;
c (7) where c (n) == if n = 0 then 1 else n * c (n - 1) endif

-- Test for recursion in functions

TESTS OF COMPLEX FUNCTIONS, INCLUDING BOOLEANS AS ARGUMENTS AND RESULTS

let c : $Z ->$B;
c (1) where
c (n) == n = 6

-- Test for booleans in function

let c : $Z * $Z * $Z ->$Z;
c(2 - 1,3,4) where c(n,m,x) == n * m * x

-- Test for multiple arguments

let c : $Z ->$B;
let d : $Z ->$Z;
c (1) where
c (n) == 1 = d(1) where
d(k) == k

-- Test for chaining in functions

let c : $Z ->$Z;
let d : $Z ->$Z;
let e : $Z ->$B;
c (3) where
c (n) == 1 + d(n) where
d(k) == if e(1) then k else k + 1 endif
where e (k) == k = 3

-- Test for nesting in functions
let $c : \mathbb{Z} \rightarrow \mathbb{Z}$;
let $d : \mathbb{Z} \rightarrow \mathbb{Z}$;
let $e : \mathbb{Z} \rightarrow \mathbb{B}$;

$c (3) \times 10$ where
$c (n) == 1 + d(n)$ where
$d(k) == \text{if } e(1) \text{ then } k \text{ else } k + 1 \text{ endif}
\text{ where } e(k) == k = 3$

-- Test for nesting in functions, result multiplied by constant

let $c : \mathbb{Z} \rightarrow \mathbb{Z}$;
let $d : \mathbb{Z} \rightarrow \mathbb{Z}$;
let $e : \mathbb{Z} \rightarrow \mathbb{B}$;

$c(3) \times c(4)$ where
$c(n) == 1 + d(n)$ where
$d(k) == \text{if } e(2 = 3 \land 4 = 5) \text{ then } k \text{ else } k + 1 \text{ endif}
\text{ where } e(k) == k$

and $b == 10$

-- Test for two functions, same definition
-- Also, test for extraneous variable defined at end of program

let $c : \mathbb{Z} \rightarrow \mathbb{Z}$;
let $d : \mathbb{Z} \rightarrow \mathbb{Z}$;
let $e : \mathbb{B} \rightarrow \mathbb{B}$;

$c(3) \times c(4)$ where
$c(n) == 1 + d(n)$ where
$d(k) == \text{if } e(2 = 3 \land 4 = 5) \text{ then } k \text{ else } k + 1 \text{ endif}
\text{ where } e(k) == k$

-- Test for boolean expression as an actual

TESTS FOR "AND" AND "WHERE" NESTING AND COMBINATIONS

let $c : \mathbb{Z} \rightarrow \mathbb{Z}$;
let $d : \mathbb{Z} \rightarrow \mathbb{Z}$;

$c(3) \times b$ where $b == 10$ and
c(n) == $n \times d(n)$ where
d(n) == 3

-- Test for nesting in functions
let c : \$Z \rightarrow \$Z;
let d : \$Z \rightarrow \$Z;

\[ c (3) \ast b \text{ where } b == 10 \text{ and } \]
\[ c (n) == n \ast d (n) \text{ where } \]
\[ d (n) == 3 \ast e \text{ where } e == 10 \]

-- Test for nesting in functions

let c : \$Z \rightarrow \$Z;
let d : \$Z \rightarrow \$Z;
let e : \$Z \rightarrow \$Z;

\[ c (3) + b \text{ where } b == 10 \text{ and } \]
\[ c (n) == d (1) + \text{ if } n == e (1) \text{ then } 2 \text{ else } 10 \text{ endif } \]
\[ \text{ where } e (k) == -1 \text{ and } \]
\[ d (g) == g + 5 \]

-- Test for nested wheres and ands

let c : \$Z \rightarrow \$Z;
let d : \$Z \rightarrow \$Z;
let e : \$Z \rightarrow \$B;

\[ c (3) \text{ where } \]
\[ c (n) == 1 + d(n) \text{ where } \]
\[ d(k) == \text{ if } e(1) \text{ then } k \text{ else } k + 1 \text{ endif } \]
\[ \text{ where } e (b) == b == 3 \]

-- Test for nesting in functions

let c : \$Z \rightarrow \$Z;
let d : \$Z;

\[ c(5) \text{ where } c (n) == d \]
\[ \text{ and } d == 10 \ast 5 \]

-- Test for single and statement

-- Test for datadef declaration

let c : \$Z;
let d : \$Z;
let e : \$Z;

c where c == (d + 10 + e \text{ where } e == 10)
and $d == 10$

-- Test for Multiple ands

let $c : \mathbb{Z}$;
let $d : \mathbb{Z}$;
let $e : \mathbb{Z}$;

$c$ where $c == d + 10 + e$
  and $d == 10$
  and $e == 10$

-- Test for Multiple ands

let $c : \mathbb{Z} \rightarrow \mathbb{Z}$;
let $d : \mathbb{Z} \rightarrow \mathbb{Z}$;
let $e : \mathbb{Z} \rightarrow \mathbb{Z}$;

c(5) where $c(n) == d(n) + 12$
  and $d(s) == 10 + s$

-- Test for Multiple ands using functions

let $c : \mathbb{Z} \rightarrow \mathbb{Z}$;
let $d : \mathbb{Z} \rightarrow \mathbb{Z}$;
let $e : \mathbb{Z} \rightarrow \mathbb{Z}$;

c(5) where $c(n) == d(n) + 12$
  and $d(s) == 10 + e(s)$
  and $e(k) == 20 + k + t$ where $t == 100$

-- Test for Multiple ands, nested wheres

let $c : \mathbb{Z}$;
let $d : \mathbb{Z}$;
let $e : \mathbb{Z}$;

$c$ where $c == d + 10 + e$ where
  $e == 10$ and $d == 10$

-- Test for Multiple ands
\( c(1) \land d(2) \) where
\[
\begin{align*}
c(n) &= n = 3 \text{ and} \\
d(n) &= (1 = k(n - 1)) \text{ where} \\
k(l) &= l + 10)
\end{align*}
\]

-- Test for proper use of "and" and implementation of
-- Paren

let \( c : \mathbb{Z} \rightarrow \mathbb{Z} \);
let \( d : \mathbb{Z} \rightarrow \mathbb{Z} \);
let \( e : \mathbb{Z} \rightarrow \mathbb{Z} \);
\( c(5) \) where \( c(n) = d(n) + 12 \) where \( k = 100 \)
and \( d(s) = 10 + e(s) \)
and \( e(k) = 20 + k \)

-- Test for Multiple ands, multiple wheres and formal/variable collisions

let \( c : \mathbb{Z} \rightarrow \mathbb{Z} \);
let \( d : \mathbb{Z} \rightarrow \mathbb{Z} \);
let \( e : \mathbb{Z} \rightarrow \mathbb{Z} \);
\( c(5) \) where \( c(n) = d(n) + 12 \) where \( k = 100 \)
and \( d(s) = 10 + e(s) \) where \( t = 100 \)
and \( e(k) = 20 + k + t \)

-- Test for Multiple ands, multiple wheres and formal/variable collisions

let \( c : \mathbb{Z} \rightarrow \mathbb{Z} \);
let \( d : \mathbb{Z} \rightarrow \mathbb{Z} \);
let \( e : \mathbb{Z} \rightarrow \mathbb{Z} \);
\( c(5) \) where \( c(n) = d(n) + 12 \) where \( t = 100 \)
and \( d(s) = 10 + e(s) \) where \( t = 120 \)
and \( e(k) = 20 + k + t \)

-- Test for Multiple ands, multiple wheres and formal/variable collisions
-- Also test to see if the proper "t" (120) was picked up

let \( c : \mathbb{Z} * \mathbb{Z} \rightarrow \mathbb{Z} \);
let \( d : \mathbb{Z} * \mathbb{Z} \rightarrow \mathbb{Z} \);
let \( e : \mathbb{Z} * \mathbb{Z} \rightarrow \mathbb{Z} \);
c(5,1) where c(n,m) == d(n,m) + 12 where t == 100
and d(s,z) == 10 + e(s,z) where t == 120
and e(k,l) == 20 + k + t + l

-- Test for Multiple ands, multiple wheres and formal/variable collisions
-- Test specifically for functions with multiple arguments

let c : $Z -> $Z;
let d : $Z -> $Z;
let e : $Z -> $Z;
c(5) where c(n) == d(n) where t == 100
and d(s) == e(s) where k == 2)
and e(k) == 20 + t

-- Test for Multiple ands, multiple wheres and formal/variable collisions

let c : $Z -> $Z;
let d : $Z -> $Z;
let e : $Z -> $Z;
c(10) where c(n) == d(n) where t == 100
and d(s) == e(s) where k == 10
and e(r) == 20 + r + k

-- Test for Multiple ands, multiple wheres and formal/variable collisions

let c : $Z -> $Z;
let d : $Z -> $Z;
let e : $Z -> $Z;
c(10) where c(n) == d(n) + t where t == (r * 100 where r == 2)
and d(s) == e(s) where k == 10
and e(r) == 20 + r + k

-- Test for Multiple ands, multiple wheres and formal/variable collisions

let c : $Z -> $Z;
let d : $Z -> $Z;
let e : $Z -> $Z;
let f : $N -> $Z;
c(10) where c(n) == d(n) + t where t == (r * 100 where r == 2)
and d(s) == e(s) where k == 10
and e(r) == 20 + r + f(r)
and f(r) == r

-- Test for Multiple ands, multiple wheres and formal/variable collisions

let c : $Z -> $Z;
let d : $Z -> $Z;
let e : $Z -> $Z;
let f : $N -> $Z;

c(10) where c(n) == d(n) + t where t == (r * 100 where r == 2)
    and d(s) == e (s) where k == 10
    and e(r) == 20 + r + f(r)
    and f(r) == k

-- Test for Multiple ands, multiple wheres and formal/variable collisions

let c : $Z -> $Z;
let d : $Z -> $Z;
let e : $Z -> $Z;
let f : $N -> $Z;

c(10) where c(n) == d(n) + t where t == (r * 100 where r == 2)
    and d(s) == e (s) where k == 10
    and e(r) == 20 + r + f(r)
    and f(r) == if r == 0 then 100 else f (r - 1) endif

-- Test for Multiple ands, multiple wheres and formal/variable collisions
-- Test for if-then-else collisions with multiple ands, wheres

let c : $Z -> $Z;
let d : $Z -> $Z;
let e : $Z -> $Z;
let f : $N -> $Z;
let zebra : $Z;

c(10) where c(n) == d(n) + t where t == (r * 100 where r == 2)
    and d(s) == e (s) where k == 10
    and e(r) == 20 + r + f(r) + zebra
    and f(r) == if r == 0 then 100 else f (r - 1) endif
    and zebra == t)

-- Test for Multiple ands, multiple wheres and formal/variable collisions
-- Test for if-then-else collisions with multiple ands, wheres

let c : $Z -> $Z;
let d : $Z -> $Z;
let e : $Z -> $Z;
\(c(5) \text{ where } c(n) = d(n) + 12 \text{ where } t = 100\)
and \(d(s) = (10 + e(s)) \text{ where } k = 100\)
and \(e(k) = 20 + k + t\)

-- Note the use of parenthesis here: if they are removed, the program will
-- bomb because \(t\) will be undefined

**ERROR TESTING**

let \(x : \text{Sz};\)
let \(j : \text{SZ};\)
let \(i : \text{Sz};\)

\(i \text{ where } t = x \% j\)
and \(x = 5 \text{ and } j = 0\)

-- Gives Division by Zero run time error

let \(b : \text{Sb};\)
let \(i : \text{SZ};\)
let \(j : \text{Sz};\)
let \(n : \text{Sn};\)
let \(x : \text{Sz};\)

if \(b\) then \(i\)
elsif \(-b \& b\) then \(j\)
else \(x\) endif

\(b = i = 2\) where
\(j = 0\)
and \(j\)
and \(z = 69\)

-- Gives two parser errors: line 13 and 14, \(j\) undefined and
-- where following "and"

let \(\text{fac} : \text{SN} > \text{SN};\)

\(\text{fac} (5) \text{ where } \text{fac} (n) = \text{fac} (n - 1)\)

-- Check for stack overflow

\(\text{too much} \text{ where } \text{too much} = 10000 \ast 10000\)

Check for Multiplication Overflow

\(\text{too much} \text{ where } \text{too much} = 40000 \ast 40000\)

Check for Addition overflow
too_much where too_much == -30000 - 30000

-- Check for Subtraction Overflow

let c : $Z -> $B;
let d : $Z -> $B;
let k : $Z -> $Z;
let g : $Z -> $Z;

c(l) = d(2) where
d(n) = (1 = k(n - 1)) where
k(1) = 1 + 10 and
c(n) = n = 3

-- Test for proper use of comments; note that there is no
delimiter on the second line of comments, as there should
-- be

MISCELLANEOUS TESTS

let b : $b;
let i : $Z;
let j : $Z;
let n : $n;
let x : $Z;

if (b V =b) then i
elsif (b V =b) then j
else x endif where
b == 1=2 where
i ==
and j == 2
and x == 69
Test for not construct boolean constructs

let b : $b;
let i : $Z;
let j : $Z;
let n : $n;
let x : $Z;

if (b V =b) then i
elsif (b ^ =b) then j
else x endif where
b == 1=2 where
i ==
and j == 2
and x == 69

-- should give 2
-- Check and, or, not and, not or
-- Check if, else, elseif
-- Especially, check all in combination

let a:SZ;
let b:$z;
let y:$n;
let x: $z;
let f: $n*$n->$n;
let times : $n*$n->$n;

f(30,30) where
  f(a,b) == times(a,b) where
    times(x,y) == x*y
-- Multiargument Checking
-- Natural Type Checking

let a:SZ;
let b:$z;
let y:$z;
let x: $z;
let f: $z*$z->$z;
let times : $n*$n->$z;

f(30,4) where
  f(a,b) == times(a,b) where
    times(x,y) ==
      if ( 1 == 1 ) then x%y
      else 2 endif end
-- Integer Division Checking

let c : $Z -> $B;
let d : $Z -> $B;
let k : $Z -> $Z;
let g : $Z -> $Z.

c(1) \land d(2) where
  d ( n ) == ( 1 = k ( n - 1 ) ) where
    k ( l ) == 1 + 10 ) and
  c ( n ) == n = 3

-- Test for proper use of "and" and implementation of
-- Parens
APPENDIX O - ROCK COMPILER USER'S MANUAL

I. Installation

The rock compiler program comes on a 5.25" disk with all public domain programs necessary to run it. To install this program on another floppy disk or a hard disk, use the following procedures:

1) Change the system drive to the disk drive containing the floppy disk.

2) Type "INSTALL", followed by a space and the drive and directory on which you want the program installed.

Note that the Rock compiler uses three unsupplied files to operate: RASM86, LINK86, and your choice of word processor. The RASM86 and LINK86 programs must be installed on the same directory as the compiler.

II. Running the Compiler

a. Type in "ROCK" and wait for the screen display shown in figure 1 to appear.

   ROCK COMPILER
   Press Escape Key to Exit Compiler

   Program to Compile ->

   Figure 1

b. When the prompt appears, type in the file name of the source file you want to compile, then press return. The compiler will accept directory specifications in the file designation. If the source file is found, the compilation will begin immediately, and the screen will appear as shown in figure 2. If the file is not found, the screen will appear as shown in figure 3.
c. If a successful compilation takes place, the prompt for a source file reappears. If the compilation is not successful, error messages will appear on the screen, and a copy of these messages can be found in a file named Errors.Phi. A typical error display is shown in figure 4. After perusing the errors, you may press any key to return to the prompt for a source file.
ROCKY ERRORS

line 1: formals list missing or error in formals list
line 1: misplaced or missing ==
number of errors = 2

PRESS ANY KEY TO CONTINUE

Figure 4

d. If compilation is successful, both an .exe and an .obj file will be created. In the event that an error occurs, neither file will be created.
WARNING: If you choose to compile two programs with the same prefix, ensure you save the first one before compiling the second one. Otherwise, the second compilation will overwrite the output file of the first compilation.

e. To cleanly stop the compiler, press the ESCAPE key any time the system asks for an input. If you have started to compile a program and you need a "panic" exit, press "Control-Break". If you do this, the cursor will not reappear on the screen. However, you can get it back by running the ROCK program again and making a normal exit.

III. Error Handling

Errors are divided into two categories—those found during compilation and those found during run time. The following two sections list the errors messages from both categories which you might encounter. Each message includes a brief synopsis of what causes the error.

COMPILER ERRORS

<table>
<thead>
<tr>
<th>Message</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>incomplete \ =&gt;</td>
<td>Either an I or \ was found where \ =&gt; was expected</td>
</tr>
<tr>
<td>without following /</td>
<td>A single backslash was found where a logical or construct (/) was expected</td>
</tr>
</tbody>
</table>

logical OR is \V
An incomplete type declaration was found.

An illegal constant was found; in this example, "3."

An unterminated literal was found, or a literal spanned more than one line.

A character with no meaning was found in the source file; '#', in this example.

The source program is too big for the host machine to compile.

An illegal statement follows the specified character: '*', in the example.

An illegal type definition follows the specified character: '*', in the example.

An unspecified error was found after LET, and the compiler is so completely sandbagged that it cannot recover.

A declaration, preceded by "LET", was not followed by a semicolon.

An invalid expression was found.

An expression defining a type was either missing or incorrect.

Formals were expected but not found, or formals were incompletely specified.

A PHI keyword or delimiter was expected or not found; ')', in the example.

TYPE found without an identifier.

Improper or no expression found following AND.
<table>
<thead>
<tr>
<th>Error</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>missing or invalid auxdef</td>
<td>Improper or no definition following WHERE</td>
</tr>
<tr>
<td>after keyword WHERE</td>
<td></td>
</tr>
<tr>
<td>missing or misplaced closing paren in formals list</td>
<td>Formals found without closing parenthesis.</td>
</tr>
<tr>
<td>error in processing multiple Actuals</td>
<td>One actual was found, but an error was spotted in a subsequent actual.</td>
</tr>
<tr>
<td>missing literal after keyword FILE</td>
<td>FILE was found without a filename being designated.</td>
</tr>
<tr>
<td>missing or invalid exp following KEYWORD</td>
<td>A keyword was spotted, but the following expression was illegal.</td>
</tr>
<tr>
<td>IF statement w/o ENDIF</td>
<td>No ENDIF to close off an IF statement.</td>
</tr>
<tr>
<td>error in formals preceding &quot;!-&gt;&quot;</td>
<td>&quot;!-&gt;&quot; found, but the formals list preceding it contained an error.</td>
</tr>
<tr>
<td>missing or invalid QualExp following COMMA operator</td>
<td>A list of elements was found with an illegal expression in it.</td>
</tr>
<tr>
<td>error in ArgBinding - check QualExp or closing bracket</td>
<td>An improper expression in an argument binding was found, or the closing bracket on an argument binding was not found.</td>
</tr>
<tr>
<td>OZONE LEVEL I</td>
<td>Unimplemented feature found. for 19.99 the feature can be implemented in 1999</td>
</tr>
<tr>
<td>NUMERIC VALUE EXPECTED</td>
<td>Non-numeric type found where a numeric type was expected.</td>
</tr>
<tr>
<td>NATURAL EXPECTED</td>
<td>Natural type was not found where it was expected.</td>
</tr>
<tr>
<td>INTEGER OR NATURAL EXPECTED</td>
<td>Either an integer or natural type is proper, but neither was found</td>
</tr>
<tr>
<td>ERROR IN TUPLE DEFINITION</td>
<td>A tuple is improperly defined the source file used improper types or number of types in defining the tuple. This can also mean a single variable was improperly defined</td>
</tr>
<tr>
<td>UNDEFINED VARIABLE IN AND SCOPE</td>
<td>An undefined variable was found in one of the two branches of an</td>
</tr>
</tbody>
</table>
FUNCTION WITHOUT FUNCTION DEFINITION
A function was defined without a declaration of its type and formals.

FORMALS MISMATCHED
Formals in a function definition are not the same in either type or number as those in the function's declaration.

FUNCTION CALLED WITHOUT FUNCTION DEFINITION
No function definition found for the function called.

REAL NUMBER EXPECTED
An incorrect type was found where a real number was expected.

INVALID CONSTANT EXPRESSION
An invalid constant was found.

BOOLEAN VALUE EXPECTED
A boolean value was expected, but none was found.

BOOLEAN OPERATOR EXPECTED
A boolean operator was expected, but none was found.

OUT OF RUN-TIME MEMORY SPACE
Not enough space to accommodate the program during run-time.

RUN-TIME ERRORS

DIVISION BY ZERO
Division by zero attempted.

MULTIPLICATION OVERFLOW
A multiplication operation resulted in a numeric value outside the language limits.

ADDITION OVERFLOW
An addition operation resulted in a numeric value outside the language limits.

SUBTRACTION OVERFLOW
A subtraction operation resulted in a numeric value outside the language limits.

STACK/variable SPACE CRASH
The stack overwrote the variable space.
# INITIAL DISTRIBUTION LIST

<table>
<thead>
<tr>
<th>No.</th>
<th>Copies</th>
<th>Address</th>
</tr>
</thead>
</table>
| 1.  | 2      | Library, Code Ø142  
Naval Postgraduate School  
Monterey, California 93943-5002 |
| 2.  | 2      | Chairman, Code 52  
Department of Computer Science  
Naval Postgraduate School  
Monterey, California 93943 |
| 3.  | 5      | Computer Technology Programs, Code 37  
Naval Postgraduate School  
Monterey, California 93943 |
| 4.  | 1      | D. L. Davis, Code 52Dv  
Department of Computer Science  
Naval Postgraduate School  
Monterey, California 93943 |
| 5.  | 1      | B. J. MacLennan, Code 52Mi  
Department of Computer Science  
Naval Postgraduate School  
Monterey, California 93943 |
| 6.  | 3      | Capt. J. E. Connell, USMC  
#5 Jack Ray Park  
Wentzville, Missouri 63385 |
| 7.  | 3      | Maj. E. J. Cole, USMC  
156 Haviland Rd  
Ridgefield, Connecticut Ø6877 |
| 8.  | 2      | Defense Technical Information Center  
Cameron Station  
Alexandria, Virginia 223Ø4-6145 |
END
9-87
DTIC