Preface

Changes for the Tenth Edition

The goals, overall structure, and approach of this tenth edition of *Concepts of Programming Languages* remain the same as those of the nine earlier editions. The principal goals are to introduce the main constructs of contemporary programming languages and to provide the reader with the tools necessary for the critical evaluation of existing and future programming languages. A secondary goal is to prepare the reader for the study of compiler design, by providing an in-depth discussion of programming language structures, presenting a formal method of describing syntax and introducing approaches to lexical and syntatic analysis.

The tenth edition evolved from the ninth through several different kinds of changes. To maintain the currency of the material, some of the discussion of older programming languages has been removed. For example, the description of COBOL's record operations was removed from Chapter 6 and that of Fortran's `Do` statement was removed from Chapter 8. Likewise, the description of Ada's generic subprograms was removed from Chapter 9 and the discussion of Ada's asynchronous message passing was removed from Chapter 13.

On the other hand, a section on closures, a section on calling subprograms indirectly, and a section on generic functions in F# were added to Chapter 9; sections on Objective-C were added to Chapters 11 and 12; a section on concurrency in functional programming languages was added to Chapter 13; a section on C# event handling was added to Chapter 14;. a section on F# and a section on support for functional programming in primarily imperative languages were added to Chapter 15.

In some cases, material has been moved. For example, several different discussions of constructs in functional programming languages were moved from Chapter 15 to earlier chapters. Among these were the descriptions of the control statements in functional programming languages to Chapter 8 and the lists and list operations of Scheme and ML to Chapter 6. These moves indicate a significant shift in the philosophy of the book—in a sense, the mainstreaming of some of the constructs of functional programming languages. In previous editions, all discussions of

functional programming language constructs were segregated in Chapter 15.

Chapters 11, 12, and 15 were substantially revised, with five figures being added to Chapter 12.

Finally, numerous minor changes were made to a large number of sections of the book, primarily to improve clarity.

The Vision

This book describes the fundamental concepts of programming languages by discussing the design issues of the various language constructs, examining the design choices for these constructs in some of the most common languages, and critically comparing design alternatives.

Any serious study of programming languages requires an examination of some related topics, among which are formal methods of describing the syntax and semantics of programming languages, which are covered in Chapter 3. Also, implementation techniques for various language constructs must be considered: Lexical and syntax analysis are discussed in Chapter 4, and implementation of subprogram linkage is covered in Chapter 10. Implementation of some other language constructs is discussed in various other parts of the book.

The following paragraphs outline the contents of the ninth edition.

Chapter Outlines

Chapter 1 begins with a rationale for studying programming languages. It then discusses the criteria used for evaluating programming languages and language constructs. The primary influences on language design, common design trade-offs, and the basic approaches to implementation are also examined.

Chapter 2 outlines the evolution of most of the important languages discussed in this book. Although no language is described completely, the origins, purposes, and contributions of each are discussed. This historical overview is valuable, because it provides the background necessary to understanding the practical and theoretical basis for contemporary language design. It also motivates further study of language design and evaluation. In addition, because none of the remainder of the book depends on Chapter 2, it can be read on its own, independent of the other chapters.

Chapter 3 describes the primary formal method for describing the syntax of programming language—BNF. This is followed by a description of attribute grammars, which describe both the syntax and static semantics of languages. The difficult task of semantic description is then explored,

including brief introductions to the three most common methods: operational, denotational, and axiomatic semantics.

Chapter 4 introduces lexical and syntax analysis. This chapter is targeted to those colleges that no longer require a compiler design course in their curricula. Like Chapter 2, this chapter stands alone and can be read independently of the rest of the book.

Chapters 5 through 14 describe in detail the design issues for the primary constructs of programming languages. In each case, the design choices for several example languages are presented and evaluated. Specifically, Chapter 5 covers the many characteristics of variables, Chapter 6 covers data types, and Chapter 7 explains expressions and assignment statements. Chapter 8 describes control statements, and Chapters 9 and 10 discuss subprograms and their implementation. Chapter 11 examines data abstraction facilities. Chapter 12 provides an in-depth discussion of language features that support object-oriented programming (inheritance and dynamic method binding), Chapter 13 discusses concurrent program units, and Chapter 14 is about exception handling, along with a brief discussion of event handling.

The last two chapters (15 and 16) describe two of the most important alternative programming paradigms: functional programming and logic programming. However, some of the data structures and control constructs of functional programming languages are discussed in Chapters 6 and 8. Chapter 15 presents an introduction to Scheme, including descriptions of some of its primitive functions, special forms, and functional forms, as well as some examples of simple functions written in Scheme. Brief introductions to ML, Haskell, and F# are given to illustrate some different directions in functional language design. Chapter 16 introduces logic programming and the logic programming language, Prolog.

To the Instructor

In the junior-level programming language course at the University of Colorado at Colorado Springs, the book is used as follows: We typically cover Chapters 1 and 3 in detail, and though students find it interesting and beneficial reading, Chapter 2 receives little lecture time due to its lack of hard technical content. Because no material in subsequent chapters depends on Chapter 2, as noted earlier, it can be skipped entirely, and because we require a course in compiler design, Chapter 4 is not covered.

Chapters 5 through 9 should be relatively easy for students with extensive programming experience in C++, Java, or C#. Chapters 10 through 14 are more challenging and require more detailed lectures.

Chapters 15 and 16 are entirely new to most students at the junior level. Ideally, language processors for Scheme and Prolog should be available for students required to learn the material in these chapters.

Sufficient material is included to allow students to dabble with some simple programs.

Undergraduate courses will probably not be able to cover all of the material in the last two chapters. Graduate courses, however, should be able to completely discuss the material in those chapters by skipping over parts of the early chapters on imperative languages.

Supplemental Materials

The following supplements are available to all readers of this book at *www.aw.com/cssupport*.

- A set of lecture note slides. PowerPoint slides are available for each chapter in the book.
- PowerPoint slides containing all the figures in the book.

To reinforce learning in the classroom, to assist with the hands-on lab component of this course, and/or to facilitate students in a distance-learning situation, access the companion Web site at *www.aw.com/sebesta*. This site contains mini-manuals (approximately 100-page tutorials) on a handful of languages. These proceed on the assumption that the student knows how to program in some other language, giving the student enough information to complete the chapter materials in each language. Currently the site includes manuals for C++, C, Java, and Smalltalk.

Language Processor Availability

Processors for and information about some of the programming languages discussed in this book can be found at the following Web sites:

C, C++, Fortran, and Ada	*gcc.gnu.org*
C# and F#	*microsoft.com*
Java	*java.sun.com*
Haskell	*haskell.org*
Lua	*www.lua.org*
Scheme	*www.plt-scheme.org/software/drscheme*
Perl	*www.perl.com*
Python	*www.python.org*
Ruby	*www.ruby-lang.org*

JavaScript is included in virtually all browsers; PHP is included in virtually all Web servers.

All this information is also included on the companion Web site.

Acknowledgments

The suggestions from outstanding reviewers contributed greatly to this book's present form. In alphabetical order, they are:

I-ping Chu	*DePaul University*
Amer Diwan	*University of Colorado*
Stephen Edwards	*Virginia Tech*
Nigel Gwee	*Southern University–Baton Rouge*
K. N. King	*Georgia State University*
Donald Kraft	*Louisiana State University*
Simon H. Lin	*California State University–Northridge*
Mark Llewellyn	*University of Central Florida*
Bruce R. Maxim	*University of Michigan–Dearborn*
Gloria Melara	*California State University–Northridge*
Frank J. Mitropoulos	*Nova Southeastern University*
Euripides Montagne	*University of Central Florida*
Bob Neufeld	*Wichita State University*
Amar Raheja	*California State Polytechnic University–Pomona*
Hossein Saiedian	*University of Kansas*
Neelam Soundarajan	*Ohio State University*
Paul Tymann	*Rochester Institute of Technology*
Cristian Videira Lopes	*University of California–Irvine*
Salih Yurttas	Texas A&M University

Numerous other people provided input for the previous editions of Concepts of Programming Languages at various stages of its development. All of their comments were useful and greatly appreciated. In alphabetical order, they are: Vicki Allan, Henry Bauer, Carter Bays, Manuel E. Bermudez, Peter Brouwer, Margaret Burnett, Paosheng Chang, Liang Cheng, John Crenshaw, Charles Dana, Barbara Ann Griem, Mary Lou Haag, John V. Harrison, Eileen Head, Ralph C. Hilzer, Eric Joanis, Leon Jololian, Hikyoo Koh, Jiang B. Liu, Meiliu Lu, Jon Mauney, Robert McCoard, Dennis L. Mumaugh, Michael G. Murphy, Andrew Oldroyd, Young Park, Rebecca Parsons, Steve J. Phelps, Jeffery Popyack, Raghvinder Sangwan, Steven Rapkin, Hamilton Richard, Tom Sager, Joseph Schell, Sibylle Schupp, Mary Louise Soffa, Neelam Soundarajan, Ryan Stansifer, Steve Stevenson, Virginia Teller, Yang Wang, John M. Weiss, Franck Xia, and Salih Yurnas.

Matt Goldstein, editor; Chelsea Bell, editorial assistant; and Meredith Gertz, senior production supervisor of Addison-Wesley, and Gillian Hall

of The Aardvark Group Publishing Services, all deserve my gratitude for their efforts to produce the tenth edition both quickly and carefully.

About the Author

Robert Sebesta is an Associate Professor Emeritus in the Computer Science Department at the University of Colorado–Colorado Springs. Professor Sebesta received a BS in applied mathematics from the University of Colorado in Boulder and MS and PhD degrees in computer science from Pennsylvania State University. He has taught computer science for more than 38 years. His professional interests are the design and evaluation of programming languages.

Contents

Answers to Selected Problems

Chapter 1

Problem Set:

3. Some arguments for having a single language for all programming domains are: It would dramatically cut the costs of programming training and compiler purchase and maintenance; it would simplify programmer recruiting and justify the development of numerous language dependent software development aids.

4. Some arguments against having a single language for all programming domains are: The language would necessarily be huge and complex; compilers would be expensive and costly to maintain; the language would probably not be very good for any programming domain, either in compiler efficiency or in the efficiency of the code it generated. More importantly, it would not be easy to use, because regardless of the application area, the language would include many unnecessary and confusing features and constructs (those meant for other application areas). Different users would learn different subsets, making maintenance difficult.

5. One possibility is wordiness. In some languages, a great deal of text is required for even simple complete programs. For example, COBOL is a very wordy language. In Ada, programs require a lot of duplication of declarations. Wordiness is usually considered a disadvantage, because it slows program creation, takes more file space for the source programs, and can cause programs to be more difficult to read.

7. The argument for using the right brace to close all compounds is simplicity—a right brace always terminates a compound. The argument against it is that when you see a right brace in a program, the location of its matching left brace is not always obvious, in part because all multiple-statement control constructs end with a right brace.

8. The reasons why a language would distinguish between uppercase and lowercase in its identifiers are: (1) So that variable identifiers may look different than identifiers that are names for constants, such as the convention of using uppercase for constant names and using lowercase for variable names in C, and (2) so that catenated words as names can have their first letter distinguished, as in `TotalWords`. (Some think it is better to include a connector, such as underscore.) The primary reason why a language would not distinguish between uppercase and lowercase in identifiers is it makes programs less readable, because words that look very similar are actually completely different, such as `SUM` and `Sum`.

10. One of the main arguments is that regardless of the cost of hardware, it is not free. Why write a program that executes slower than is necessary. Furthermore, the difference between a well-written efficient program and one that is poorly written can be a factor of two or three. In many other fields of endeavor, the difference between a good job and a poor job may be 10 or 20 percent. In programming, the difference is much greater.

15. The use of type declaration statements for simple scalar variables may have very little effect on the readability of programs. If a language has no type declarations at all, it may be an aid to readability, because regardless of where a variable is seen in the program text, its type can be determined without looking elsewhere. Unfortunately, most languages that allow implicitly declared variables also include explicit declarations. In a program in such a language, the

declaration of a variable must be found before the reader can determine the type of that variable when it is used in the program.

18. The main disadvantage of using paired delimiters for comments is that it results in diminished reliability. It is easy to inadvertently leave off the final delimiter, which extends the comment to the end of the next comment, effectively removing code from the program. The advantage of paired delimiters is that you *can* comment out areas of a program. The disadvantage of using only beginning delimiters is that they must be repeated on every line of a block of comments. This can be tedious and therefore error-prone. The advantage is that you cannot make the mistake of forgetting the closing delimiter.

Chapter 2

Problem Set:

6. Because of the simple syntax of LISP, few syntax errors occur in LISP programs. Unmatched parentheses is the most common mistake.

7. The main reason why imperative features were put in LISP was to increase its execution efficiency.

10. The main motivation for the development of PL/I was to provide a single tool for computer centers that must support both scientific and commercial applications. IBM believed that the needs of the two classes of applications were merging, at least to some degree. They felt that the simplest solution for a provider of systems, both hardware and software, was to furnish a single hardware system running a single programming language that served both scientific and commercial applications.

11. IBM was, for the most part, incorrect in its view of the future of the uses of computers, at least as far as languages are concerned. Commercial applications are nearly all done in languages that are specifically designed for them. Likewise for scientific applications. On the other hand, the IBM design of the 360 line of computers was a great success--it still dominates the area of computers between supercomputers and minicomputers. Furthermore, 360 series computers and their descendants have been widely used for both scientific and commercial applications. These applications have been done, in large part, in Fortran and COBOL.

14. The argument for typeless languages is their great flexibility for the programmer. Literally any storage location can be used to store any type value. This is useful for very low-level languages used for systems programming. The drawback is that type checking is impossible, so that it is entirely the programmer's responsibility to insure that expressions and assignments are correct.

18. A good deal of restraint must be used in revising programming languages. The greatest danger is that the revision process will continually add new features, so that the language grows more and more complex. Compounding the problem is the reluctance, because of existing software, to remove obsolete features.

22. One situation in which pure interpretation is acceptable for scripting languages is when the amount of computation is small, for which the processing time will be negligible. Another situation is when the amount of computation is relatively small and it is done in an interactive environment, where the processor is often idle because of the slow speed of human interactions.

24. New scripting languages may appear more frequently than new compiled languages because they are often smaller and simpler and focused on more narrow applications, which means their libraries need not be nearly as large.

Chapter 3

Instructor's Note:

In the program proof on page 160, there is a statement that may not be clear to all, specifically, `(n + 1)* … * n = 1`. The justification of this statement is as follows:

Consider the following expression:

```
(count + 1) * (count + 2) * … * n
```

The former expression states that when `count` is equal to `n`, the value of the later expression is `1`. Multiply the later expression by the quotient:

```
(1 * 2 * … * count) / (1 * 2 * … * count)
```

whose value is `1`, to get

```
(1 * 2 * … * count * (count + 1) * (count + 2) * … * n) /

     (1 * 2 * … * count)
```

The numerator of this expressions is `n!`. The denominator is `count!`. If `count` is equal to `n`, the value of the quotient is

```
n! / n!
```

or `1`, which is what we were trying to show.

Problem Set:

2a. <class_head> → {<modifier>} **`class`** <id> [**`extends`** class_name]

[**`implements`** <interface_name> {, <interface_name>}]

<modifier> → **`public`** | **`abstract`** | **`final`**

2c. <switch_stmt> → **`switch`** (<expr>) { **`case`** <literal> : <stmt_list>

{**`case`** <literal> : <stmt_list> } [**`default`** : <stmt_list>] }

3. <assign> → <id> = <expr>

<id> → A | B | C

<expr> → <expr> * <term>

| <term>

<term> → <factor> + <term>

| <factor>

<factor> → (<expr>)

| <id>

6.

(a) <assign> => <id> = <expr>

=> A = <expr>

=> A = <id> * <expr>

=> A = A * <expr>

=> A = A * (<expr>)

=> A = A * (<id> + <expr>)

=> A = A * (B + <expr>)

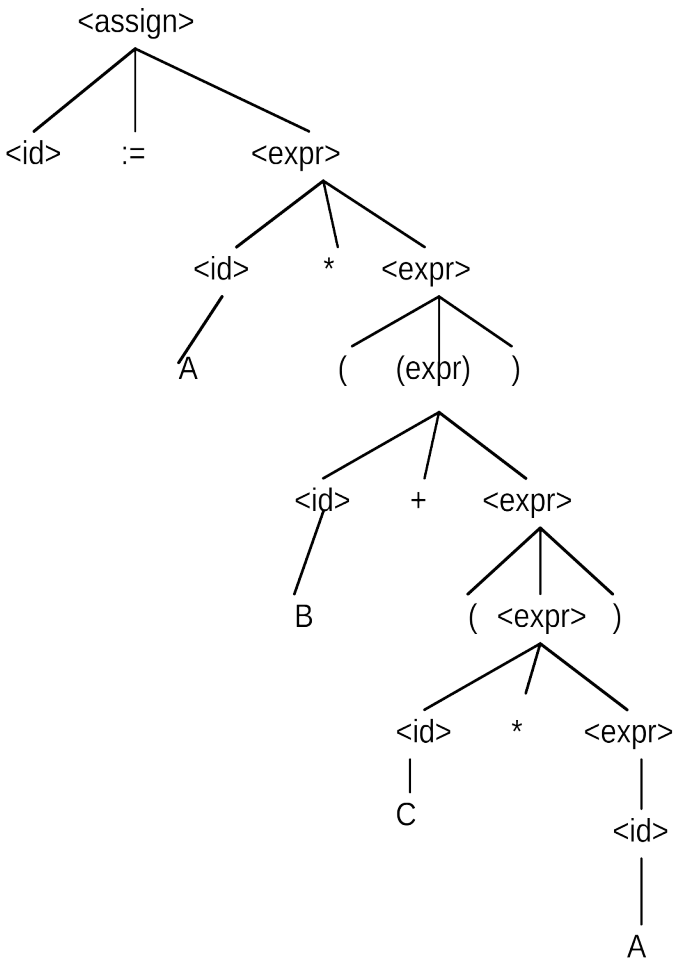
=> A = A * (B + (<expr>))

=> A = A * (B + (<id> * <expr>))

=> A = A * (B + (C * <expr>))

=> A = A * (B + (C * <id>))

=> A = A * (B + (C * A))



7.

(a) <assign> => <id> = <expr>

=> A = <expr>

=> A = <term>

=> A = <factor> * <term>

=> A = (<expr>) * <term>

=> A = (<expr> + <term>) * <term>

=> A = (<term> + <term>) * <term>

=> A = (<factor> + <term>) * <term>

=> A = (<id> + <term>) * <term>

=> A = (A + <term>) * <term>

=> A = (A + <factor>) * <term>

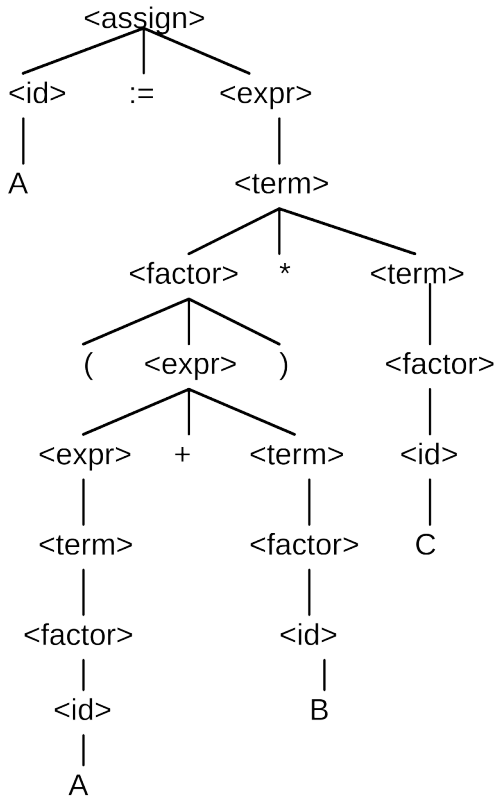
=> A = (A + <id>) * <term>

=> A = (A + B) * <term>

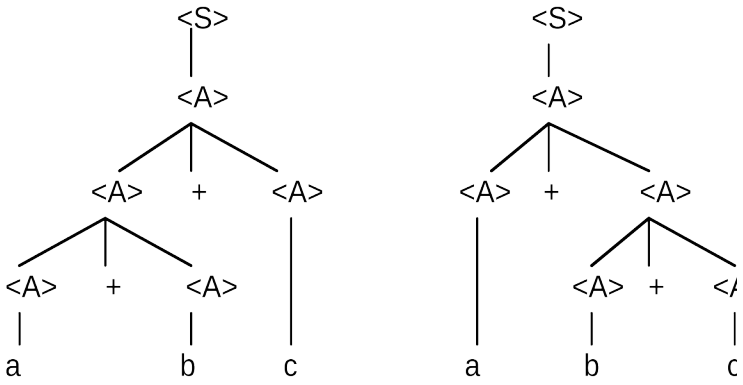
=> A = (A + B) * <factor>

=> A = (A + B) * <id>

=> A = (A + B) * C



8. The following two distinct parse tree for the same string prove that the grammar is ambiguous.



9. Assume that the unary operators can precede any operand. Replace the rule

<factor> → <id>

with

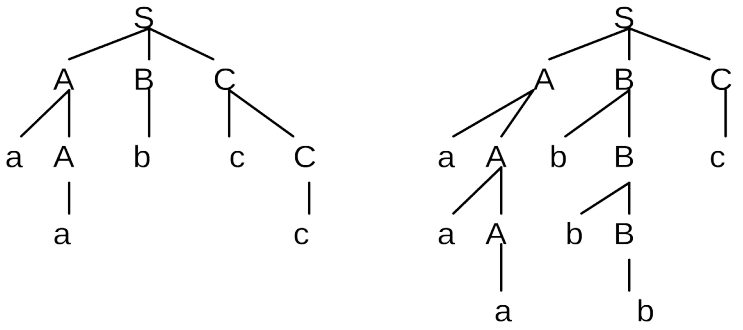
<factor> → + <id>

| - <id>

10. One or more a's followed by one or more b's followed by one or more c's.

13. S → a S b | a b

14.



16. <assign> → <id> = <expr>

<id> → A | B | C

<expr> → <expr> (+ | -) <expr>

| (<expr>)

| <id>

18. The value of an intrinsic attribute is supplied from outside the attribute evaluation process, usually from the lexical analyzer. A value of a synthesized attribute is computed by an attribute evaluation function.

19. Replace the second semantic rule with:

<var>[2].env ← <expr>.env

<var>[3].env ← <expr>.env

<expr>.actual_type ← <var>[2].actual_type

predicate: <var>[2].actual_type = <var>[3].actual_type

21.

(a) (Java **`do-while`**) We assume that the logic expression is a single relational expression.

```
loop:   (do body)
        if <relational_expression> goto out
        goto loop
out:    ...
```

(b) (Ada **`for`**) **`for`** `I` **`in`** `first .. last` **`loop`**

```
        I = first
loop:   if I < last goto out
        ...
        I = I + 1
        goto loop
```

```
out:        ...
```

(c) (Fortran Do)

```
            K = start
loop:       if K > end goto out
            ...
            K = K + step
            goto loop
out:        ...
```

(e) (C **`for`**) **`for`** (expr1; expr2; expr3) ...

```
            evaluate(expr1)
loop:       control = evaluate(expr2)
            if control == 0 goto out
            ...
            evaluate(expr3)
            goto loop
out:        ...
```

22a. M_{pf}(for var in init_expr .. final_expr loop L end loop, s) $\triangleq$

if VARMAP(i, s) = **undef** for var or some i in init_expr or final_expr
then **error**
else if M_e(init_expr, s) > M_e(final_expr, s)
then s
else M_l(while init_expr - 1 <= final_expr do L, M_a(var := init_expr + 1, s))

22b. M_r(repeat L until B) $\triangleq$

if M_b(B, s) = **undef**

then **error**

else if M_{sl}(L, s) = **error**

then **error**

else if M_b(B, s) = **true**

then M_{sl}(L, s)

else M_r(repeat L until B), M_{sl}(L, s))

22c. M_b(B, s) $\triangleq$ if VARMAP(i, s) = **undef** for some i in B

then **error**

else B', where B' is the result of

evaluating B after setting each

variable i in B to VARMAP(i, s)

22d. M_{cf}(for (expr1; expr2; expr3) L, s) $\triangleq$

if VARMAP (i, s) = **undef** for some i in expr1, expr2, expr3, or L

then **error**

else if M_e (expr2, M_e (expr1, s)) = 0

then s

else M_{help} (expr2, expr3, L, s)

M_{help} (expr2, expr3, L, s) $\triangleq$

if VARMAP (i, s) = **undef** for some i in expr2, expr3, or L

then **error**

else

if M_{sl} (L, s) = **error**

then s

else M_{help} (expr2, expr3, L, M_{sl} (L, M_e (expr3, s))

23.

(a) $a = 2 * (b - 1) - 1 \quad \{a > 0\}$

$2 * (b - 1) - 1 > 0$

$2 * b - 2 - 1 > 0$

$2 * b > 3$

$b > 3 / 2$

(b) $b = (c + 10) / 3 \ \{b > 6\}$

$(c + 10) / 3 > 6$

$c + 10 > 18$

$c > 8$

(c) $a = a + 2 * b - 1 \ \{a > 1\}$

$a + 2 * b - 1 > 1$

$2 * b > 2 - a$

$b > 1 - a / 2$

(d) $x = 2 * y + x - 1 \ \{x > 11\}$

$2 * y + x - 1 > 11$

$2 * y + x > 12$

24.

(a) $a = 2 * b + 1$

$b = a - 3 \quad \{b < 0\}$