The final example extends the calculator to make it a small but somewhat realistic “compiler”.

We’ll add
named variables and assignments;
comparison expressions (greater, less, equal, etc.);
flow control with if/then/else and while/do;
built-in and user-defined functions;
and a little error recovery.

The previous version of the calculator didn’t take much advantage of the AST representation of expressions, but in this one, the AST is the key to the implementation of flow control and user functions.
Conceptually, there are two types of expressions: those that assign a value to a variable and those that simply have a value.

The expression $x = 7$ is an example of the first type. This expression uses the $=$ operator to assign the value seven to the variable $x$. The expression itself evaluates to seven.

The code $3 + 4$ is an example of the second expression type. This expression uses the $+$ operator to add three and four together without assigning the result, seven, to a variable.
Assignment in predicate can be useful for loops more than if statements.

```c
while( var = GetNext() )
{
    ...do something with var
}
```

Which would otherwise have to be written

```c
var = GetNext();
while( var )
{
    ...do something
    var = GetNext();
}
```
Here's an example of defining a user function, and then calling it, using a built-in function as one of the arguments:

```plaintext
> let avg(a,b) = (a+b)/2;
Defined avg
> avg(3, sqrt(25))
= 4
```

The math library must be linked in when building the executable. How to do this varies by environment, but in Linux/Unix, just add `-lm` to the command:

```
gcc lex.yy.c fb3-2.tab.c fb3-2funcs.c -lm
```
let max(x,y) = if x >= y then x; else y;

> max(4+5,6+7)
In mathematics, the Euclidean algorithm, or Euclid's algorithm, is an efficient method for computing the greatest common divisor (GCD) of two numbers, the largest number that divides both of them without leaving a remainder.

It is named after the ancient Greek mathematician Euclid, who first described it in his Elements (c. 300 BC).

It is an example of an algorithm, a step-by-step procedure for performing a calculation according to well-defined rules, and is one of the oldest algorithms in common use.
let euclid(x,y) = \\
    if (x == y) then x; \\
    else if (x > y) then euclid(x-y, y); \\
    else euclid (x, y-x);\n
;;

Defined euclid

euclid(9,12)
euclid(12,20)
The six comparison operators all return a CMP token with a lexical value to distinguish them.

The six keywords and four built-in functions are recognized by literal patterns.

Note that they have to precede the general pattern to match a name so that they're matched in preference to the general pattern.
The name pattern looks up the name in the symbol table and returns a pointer to the symbol.
Advanced Calculator - HASHING

Assume that you have an object and you want to assign a key to it to make searching easy.

To store the key/value pair, you can use a simple array like a data structure where keys (integers) can be used directly as an index to store values.

However, in cases where the keys are large and cannot be used directly as an index, you should use hashing.

In hashing, large keys are converted into small keys by using hash functions.

The values are then stored in a data structure called a hash table.

The idea of hashing is to distribute entries (key/value pairs) uniformly across an array. Each element is assigned a key (converted key).

By using that key you can access the element in $O(1)$ time.

Using the key, the algorithm (hash function) computes an index that suggests where an entry can be found or inserted.
An Advanced Calculator - Symbol Table

The hash function is also quite simple: For each character, multiply the previous hash by 9 and then xor the character.

```c
/* hash a symbol */
static unsigned
symhash(char *sym)
{
    unsigned int hash = 0;
    unsigned c;

    while(c = *sym++) hash = hash*9 ^ c;

    return hash;
}
```
The lookup routine computes the symbol table entry index as the hash value modulo the size of the symbol table, which was chosen as a number with no even factors, again to mix the hash bits up.

lookup takes a string and returns the address of the table entry for that name, creating a new entry if there isn't one already. The lookup technique is known as hashing with linear probing. It uses a hash function to turn the string into an entry number in the table, then checks the entry, and, if it's already taken by a different symbol, scans linearly until it finds a free entry.

**lookup (char* sym)**

```c
struct symbol *lookup(char* sym)
{
    struct symbol *sp = &symtab[symhash(sym)%NHASH];
    int scount = NHASH;  /* how many have we looked at */

    while(--scount >= 0) {
        if(sp->name && !strcmp(sp->name, sym)) { return sp; }

        if(!sp->name) {  /* new entry */
            sp->name = strdup(sym);
            sp->value = 0;
            sp->func = NULL;
            sp->syms = NULL;
            return sp;
        }

        if(++sp >= symtab+NHASH) sp = symtab; /* try the next entry */
    }
    yyerror("symbol table overflow\n");
    abort();  /* tried them all, table is full */
}
```

**struct symbol**

```c
struct symbol {
    char *name;
    double value;
    struct ast *func;  /* stmt for the function */
    struct symlist *syms;  /* list of dummy args */
};
```

```c
#define NHASH 9997
struct symbol symtab[NHASH];
```

**struct symbol *lookup(char*)**

The lookup routine computes the symbol table entry index as the hash value modulo the size of the symbol table, which was chosen as a number with no even factors, again to mix the hash bits up.

lookup takes a string and returns the address of the table entry for that name, creating a new entry if there isn't one already. The lookup technique is known as hashing with linear probing. It uses a hash function to turn the string into an entry number in the table, then checks the entry, and, if it's already taken by a different symbol, scans linearly until it finds a free entry.

**strcmp()** returns an integer indicating the result of the comparison, as follows:

- 0, if the s1 and s2 are equal;
/* symbol table */
struct symbol { /* a variable name */
    char *name;
    double value;
    struct ast *func; /* stmt for the function */
    struct symlist *syms; /* list of dummy args */
};
There are two ways to specify precedence and associativity in a grammar, implicitly and explicitly.

So far, we’ve specified them implicitly, by using separate non-terminal symbols for each precedence level.

This is a perfectly reasonable way to write a grammar, and if bison didn’t have explicit precedence rules, it would be the only way.
The `%union` here defines many kinds of symbol values, which is typical in realistic bison parsers.

As well as a pointer to an AST and a numeric value, a value can be a pointer to the symbol table for a user symbol, a list of symbols, or a subtype of a comparison or function token.

(We use the word symbol somewhat confusingly here, both for names used in the bison grammar and for names that the user types into the compiled program.

We'll say user symbol for the latter when the context isn't otherwise clear.)
Each of these declarations defines a level of precedence, with the order of the %left, %right, and %nonassoc declarations defining the order of precedence from lowest to highest. The definition of non-associative operators is that it is illegal to combine two or more of these without explicit parentheses.
The definition of list is right recursive, that is, stmt ; list rather than list stmt ;
It doesn't make any difference to the language recognized, but it makes it easier to
build the list of statements linked from head to tail rather than from tail to head.

```
X=5;
Y=2;
z=x+y;
```

```
list: /* nothing */ { $$ = NULL; } |
  | stmt ';'; list { if ($3 == NULL)
  |     $$ = $1;
  |   else
  |     $$ = newast('L', $1, $3);
  | }
```

```
L
   
X=5;
   
L
   
Y=2;
   
L
   
z=x+y;
```

```
case 'L': eval(a->l); v = eval(a->r); break;
```
If the definition of list had
put semicolons between rather than after each statement,
the grammar would be ambiguous unless the grammar also
added closing FI and ENDDO tokens to indicate the end
of if/then and while/do statements.
The special pseudo-token error indicates an error recovery point.

Built in

Error parsing grammar go
To EOL
The rule for negation includes \texttt{%prec UMINUS}.

The only operator in this rule is \texttt{-}, which has low precedence, but we want unary minus to have higher precedence than multiplication rather than lower.

The \texttt{%prec} tells bison to use the precedence of \texttt{UMINUS} for this rule.
In the calculator, each symbol can potentially be both a variable and a user-defined function.

The value field holds the symbol’s value as a variable, the func field points to the AST for the user code for the function, and syms points to a linked list of the dummy (formal) arguments, which are themselves symbols. (In the previous example, avg is the function, and a and b are the dummy arguments.)

The C functions newsymlist and symlistfree create and free them.
Our grammar distinguishes between statements (stmt) and expressions (exp).

A statement is either a flow of control (if/then/else or while/do) or an expression.

The if and while statements take lists of statements, with each statement in the list being followed by a semicolon.

Each rule that matches a statement calls a routine to build an appropriate AST node.

```c
struct flow {
    int nodetype; /* type I or W */
    struct ast *cond; /* condition */
    struct ast *tl; /* then or do list */
    struct ast *el; /* optional else list */
};

struct ast *
newflow(int nodetype, struct ast *cond, struct ast *tl, struct ast *el)
{
    struct flow *a = malloc(sizeof(struct flow));

    if(!a) {
        yyerror("out of space");
        exit(0);
    }
    a->nodetype = nodetype;
    a->cond = cond;
    a->tl = tl;
    a->el = el;
    return (struct ast *)a;
}
```
struct symref {
  int nodetype;  /* type N */
  struct symbol *s;
};

struct symasgn {
  int nodetype;  /* type = */
  struct symbol *s;
  struct ast *v;  /* value */
};
The heart of the calculator is `eval`, which evaluates an AST built up in the parser. Following the practice in C, comparisons return 1 or 0 depending on whether the comparison succeeds, and tests in `if/then/else` and `while/do` treat any nonzero as true.
For expressions, we do the familiar depth-first tree walk to compute the value.

An AST makes it straightforward to implement if/then/else: Evaluate the condition AST to decide which branch to take, and then evaluate the AST for the path to be taken.

To evaluate while/do loops, a loop in eval evaluates the condition AST, then the body AST, repeating as long as the condition AST remains true.

Any AST that references variables that are changed by an assignment will have a new value each time it’s evaluated.
/* control flow */
/* null if/else/do expressions allowed in the grammar, so check for them */
case 'I':
  if( eval( ((struct flow *)a)->cond) != 0 ) {
    if( ((struct flow *)a)->tl ) {
      v = eval( ((struct flow *)a)->tl);
    } else
      v = 0.0; /* a default value */
  } else {
    if( ((struct flow *)a)->el ) {
      v = eval(((struct flow *)a)->el);
    } else
      v = 0.0; /* a default value */
  }
  break;

case 'W':
  v = 0.0; /* a default value */
  if( ((struct flow *)a)->tl ) {
    while( eval(((struct flow *)a)->cond) != 0 )
      v = eval(((struct flow *)a)->tl);
  }
  break; /* last value is value */

case 'L': eval(a->l); v = eval(a->r); break;

case 'F': v = callbuiltin((struct fncall *)a); break;

case 'C': v = calluser((struct ufncall *)a); break;

default: printf("internal error: bad node %s\n", a->nodetype);
} return v;
Note that if you assign a value to one specific item, the next items will update their numbers accordingly:

```c
enum bifs {
    B_sqrt = 1,
    B_exp,
    B_log,
    B_print
};
```

```c
struct fnall {
    int nodetype;
    struct ast *l;
    enum bifs functype;
};
```

```c
struct ast *
newfunc(int functype, struct ast *l)
{
    struct fnall *a = malloc(sizeof(struct fnall));

    if(!a) {
        yyerror("out of space");
        exit(0);
    }
    a->nodetype = 'F';
    a->l = l;
    a->functype = functype;
    return (struct ast *)a;
}
```
Built-in functions are relatively straightforward: Determine which function it is and call specific code to do the function.

```c
static double
callbuiltin(struct fncall *f)
{
    enum bifs functype = f->functype;
    double v = eval(f->l);
    switch(functype) {
        case B_sqrt:
            return sqrt(v);
        case B_exp:
            return exp(v);
        case B_log:
            return log(v);
        case B_print:
            printf("= %4.4g\n", v);
            return v;
        default:
            yyerror("Unknown built-in function %d", functype);
            return 0.0;
    }
}
```

```c
> sqrt(1,2,3,36)
= 6
```
A user function definition consists of the name of the function, a list of dummy arguments, and an AST that represents the body of the function.

Defining the function simply saves the argument list and AST in the function's symbol table entry, replacing any previous version.

```c
} 
| calclist LET NAME '([' symlist ']') '=>' list EOL {
  dodef($3, $5, $8);
  printf("Defined %s\n", $3->name);
}

/* define a function */
void
dodef(struct symbol *name, struct symlist *syms, struct ast *func)
{
  if(name->syms) symlistfree(name->syms);
  if(name->func) treefree(name->func);
  name->syms = syms;
  name->func = func;
}
struct ufuncall {
    int nodetype;
    struct ast *l;
    struct symbol *s;
};

struct ast *
newcall(struct symbol *s, struct ast *l)
{
    struct ufuncall *a = malloc(sizeof(struct ufuncall));

    if(!a) {
        yyerror("out of space");
        exit(0);
    }
    a->nodetype = 'C';
    a->l = l;
    a->s = s;
    return (struct ast *)a;
}
The absolute simplest example of a recursive function would be to implement multiplication by addition, the way it is defined in math:

```
let mul(a,b) = if  b == 1  then a; else  a + mul (a, b-1);;
```

```
let fib ( n ) = print(n) ; if  n < 2 then 1; else fib(n - 1) + fib(n - 2);;
```
Say you define a function to calculate the maximum of its two arguments:

```haskell
> let max(x,y) = if x >= y then x; else y;;
> max(4+5,6+7)
```

The function has two dummy arguments, \( x \) and \( y \). When the function is called, the evaluator does this:

1. Evaluate the actual arguments, \( 4+5 \) and \( 6+7 \) in this case.
2. Save the current values of the dummy arguments and assign the values of the actual arguments to them.
3. Evaluate the body of the function, which will now use the actual argument values when it refers to the dummy arguments.
4. Put back the old values of the dummies.
5. Return the value of the body expression.
static double
calluser(struct ufuncall *f)
{
    struct symbol *fn = f->s;
    int nodetype = f->n;
    struct symlist *sl;
    struct ast *args = f->a;
    double *oldval, *newval;
    double v;
    int nargs;
    int i;

    if(!fn->func) {
        yyerror("call to undefined function", fn->name);
        return 0;
    }

    /* count the arguments */
    sl = fn->symsl;
    for(nargs = 0; sl; sl = sl->next)
        nargs++;

    /* prepare to save them */
    oldval = (double *)malloc(nargs * sizeof(double));
    newval = (double *)malloc(nargs * sizeof(double));
    if(!oldval || !newval) {
        yyerror("Out of space in %s", fn->name); return 0.0;
    }

    /* evaluate the arguments */
    for(i = 0; i < nargs; i++) {
        if(!args) {
            yyerror("too few args in call to %s", fn->name);
            free(oldval); free(newval);
            return 0;
        }

        if(args->nodetype == 'L') { /* if this is a list node */
            newval[i] = eval(args->l);
            args = args->r;
        } else { /* if it's the end of the list */
            newval[i] = eval(args);
            args = NULL;
        }
    }
}
/* save old values of dummies, assign new ones */
sl = fn->syms;
for(i = 0; i < nargs; i++) {
    struct symbol *s = sl->sym;

    oldval[i] = s->value;
    s->value = newval[i];
    sl = sl->next;
}
free(newval);

/* evaluate the function */
v = eval(fn->func);

/* put the dummies back */
sl = fn->syms;
for(i = 0; i < nargs; i++) {
    struct symbol *s = sl->sym;

    s->value = oldval[i];
    sl = sl->next;
}
free(oldval);
return v;