12-1992

User Manual for Ox: An Attribute-Grammar Compiling System based on Yacc, Lex, and C

Kurt M. Bischoff

Iowa State University

Follow this and additional works at: http://lib.dr.iastate.edu/cs_techreports

Part of the Programming Languages and Compilers Commons

Recommended Citation

http://lib.dr.iastate.edu/cs_techreports/21

This Article is brought to you for free and open access by Digital Repository @ Iowa State University. It has been accepted for inclusion in Computer Science Technical Reports by an authorized administrator of Digital Repository @ Iowa State University. For more information, please contact hinefuku@iastate.edu.
User Manual for

**Ox**: An Attribute-Grammar Compiling System
based on **Yacc**, **Lex**, and **C**

by Kurt M. Bischoff

TR#92-30

December, 1992

Department of Computer Science
226 Atanasoff Hall
Iowa State University
Ames, Iowa 50011

©1992 Kurt M. Bischoff
Contents

1 Overview of Use  3
2 Preliminary  4
3 Attribute Declarations  5
  3.1 Semantics of Attribute Declarations ................. 6
4 Rules and Attribute Occurrences  6
5 Ox Attribute Definitions  7
  5.1 Inherited vs. Synthesized Attributes ............... 7
  5.2 Attribute Reference Sections in the Y-file .......... 8
    5.2.1 Explicit Mode .................................. 8
    5.2.2 Implicit Mode .................................. 10
    5.2.3 Mixed Mode .................................... 10
  5.3 Attribute Reference Sections in the L-file(s) ........ 10
  5.4 Cycles ............................................. 11
6 Translation into C Code  12
7 Temporal Behavior of the Ox-generated Evaluator  12
  7.1 Stack Operations ................................... 12
  7.2 Placement of Generated Code ......................... 13
  7.3 Decoration and the Ready Set ........................ 14
8 Programming Style  14
9 Postdecoration Traversals  15
  9.1 Example: Infix to Prefix Translation ............... 15
  9.2 General Description ................................ 18
    9.2.1 Traversal Specifications ....................... 18
    9.2.2 Traversal Action Specifications ............... 18
    9.2.3 Traversal Semantics ........................... 20
1 Overview of Use

Lex and Yacc are powerful and widely-used tools for the automatic generation of language recognizers. Lex accepts a set of user-written regular expressions and writes a C program that performs lexical analysis according to those expressions. Yacc translates user-written grammar rules into C source code for a syntax analyzer. While they afford “hooks” for execution of hand-coded C-language semantic actions, Lex and Yacc provide little other facility for automatic implementation of language semantics.

Attributed parse trees are often used as data structures in evaluators for languages. Often the language implementer hand-crafts code for building, traversing, and evaluating attributes of parse trees, and for parse-tree-related memory management. A Yacc specification defines a context-free language and a mapping from the set of legal sentences to the set of parse trees, but code for parse-tree management is not generated automatically by Yacc.

The Ox user specifies a language using the familiar languages of Lex and Yacc, or takes an existing Lex/Yacc specification, and adds semantics to the language by augmenting the specification files with declarations and definitions of typed attributes of parse-tree nodes.

That specification constitutes an attribute grammar, and from it Ox automatically generates an evaluator written in Yacc, Lex, and C. For a given input, the evaluator builds a parse tree, determines an order of evaluation for attributes of the tree, and performs, for each attribute, the semantic action required to evaluate it. This parse tree is managed independently of any trees managed by hand-written C code, but information may be moved between the evaluator-managed tree and any global data structure.

Additionally, the Ox user may easily specify parse-tree traversals that are performed after evaluation of the tree’s attributes and that refer to those attributes. Such traversals greatly simplify tasks such as code generation and the gathering of compilation statistics.

The language designer is freed from the tedious and error-prone details of writing code for parse-tree management. Ox-generated evaluators use memory-management techniques that bring large time-efficiency gains over...

\footnote{The name “Ox” originated as a homophone for an acronym for “An Attribute Grammar Compiling System”. It was noticed later that every yak is an ox and that Ox generalizes the function of Yacc.}
hand-built evaluators that use the common technique of calling `malloc` once for each parse-tree node. Also, Ox provides security by testing the definition for consistency and completeness, and the Ox-generated evaluator performs tests to ensure that a circular definition has not prevented evaluation of attributes.

Ox is a preprocessor that accepts two or more files, and translates these into files suitable for input to Lex and Yacc. With few exceptions, all Lex-input/Yacc-input pairs of files that constitute recognizers or translators are legal inputs to Ox. Thus much existing software is amenable to modification using Ox, and implementations that use Ox may be converted stepwise by hand to “pure” Lex/Yacc implementations. This makes Ox well-suited to language designers, experimenters, and implementers already familiar with Lex, Yacc, and C.

# Preliminary

It is assumed that the reader is familiar with the use of Yacc [Johnson 75], Lex [Lesk 75], and C [KR 88]; Ox syntactic constructs are described mainly as augmentations of the languages accepted by those tools. Prior acquaintance with the basic ideas of attribute grammars (for instance, as found in [Waite 84] or [Aho 86]) is helpful.

An Ox input specification consists of at least two files: a syntactic specification (which resembles a Yacc input specification and is called a Y-file) that Ox translates into a Yacc input specification, and one or more lexical specifications (which resemble Lex input specifications and are called L-files) that Ox translates into Lex input specifications. Usually there is exactly one L-file, but an evaluator that uses more than one scanner [Lesk 75] may be constructed by submitting to Ox more than one L-file. This manual presents descriptions of the Ox-specific constructs that may appear in these files, as well as pertinent underlying concepts. These constructs are illustrated mainly by using fragments of three examples of Ox input specifications, the complete texts of which appear in sections 13, 14 and 15.

Within Ox-specific constructs, C-style comments may appear anywhere whitespace may appear. The global identifiers of Ox-generated C code, like those generated by Yacc and Lex, are prefixed by `yy`, so the Ox user can avoid name conflicts in the generated evaluator by abstaining from the use
of global identifiers that begin with \texttt{yy}.

3 Attribute Declarations

As described in [Johnson 75], the declarations section of a \texttt{Yacc} input specification is the part that precedes the first \texttt{%%} mark, and in it the user may declare the start symbol, tokens, associativities, unions, C code sections, etc. The Y-file contains such a declarations section, and in it are permitted all of the constructs of \texttt{Yacc} declarations sections, as well as \texttt{Ox attribute declarations}. An attribute declaration consists of the reserved word \texttt{@attributes} followed by \{, an attribute declaration list, \}, and a list of grammar symbols.

Suppose that a grammar has a symbol \texttt{bitlist} and the following attribute declaration:

\begin{verbatim}
@attributes {float value; int scale,length;} bitlist
\end{verbatim}

Then the \texttt{Ox}-generated evaluator, when building a parse-tree node labeled \texttt{bitlist}, allocates storage for a float named \texttt{value} and integers named \texttt{scale} and \texttt{length}.

An attribute declaration list (in the previous example, the part between curly braces) resembles a C structure declaration list. Digit strings and C-style identifiers, as well as the following characters and reserved words, arranged according to C syntax, are legal in attribute declaration lists:

\begin{verbatim}
* : ; , char short int long float double
    signed unsigned struct union enum
\end{verbatim}

Note that curly braces may not appear inside (and so structures and unions may not be declared inside) attribute declaration lists. Despite this, any fundamental or derived type permitted in a C program may be used as an attribute type specifier: \texttt{Yacc} input specifications often contain C code sections between \texttt{%%} and \texttt{%%}, and these are also permitted in \texttt{Ox} input specifications. Any type name given meaning by using \texttt{struct, union, typedef,} or \texttt{#define} in a previous C code section may be used as an attribute type specifier.

The list of grammar symbols following \} is a possibly empty list of \texttt{Yacc} tokens (including character constants) and nonterminals, members of the list being separated by whitespace.
3.1 Semantics of Attribute Declarations

An attribute declaration informs Ox that each symbol in the grammar symbol list has attributes of the names and types appearing in the attribute declaration list. If \( a \) appears in the attribute declaration list and \( s \) appears in the grammar symbol list, then \( a \) is said to belong to \( s \) or to be an attribute of \( s \). Each grammar symbol has its own attribute name space. When the evaluator creates a node labeled by one of the listed symbols, it allocates storage of the specified type for each of the named attributes. A storage location so allocated is called an attribute instance (concisely, an instance) in the parse tree. Instances may be said to belong to nodes.

4 Rules and Attribute Occurrences

Yacc grammar rules (productions), and the objects of return statements in Lex actions (each such object being a token), are here referred to generically as rules. Since Ox accepts the constructs of Yacc and Lex, and passes these unchanged, the corresponding constructs of Ox input specifications are also called rules. Each rule is viewed as a sequence of grammar symbols, the object of each return statement in a Lex action being a sequence consisting of a single grammar symbol. The leftmost symbol of a rule is called the left hand side (LHS), and the right hand side (RHS) is comprised of the rule’s other symbols. A symbol’s position in a rule together with an attribute of that symbol constitute an attribute occurrence (concisely, an occurrence) in that rule. If the attribute in question is \( a \), the occurrence is said to be an occurrence of \( a \). Supposing the @attributes declaration of section 3 and the rule:

\[
\text{num} : \text{bitlist DOT bitlist}
\]

the attribute occurrence scale of the leftmost appearance of bitlist is denoted in Ox code as \text{bitlist.0.scale}, while the attribute occurrence scale of the rightmost appearance of bitlist is denoted \text{bitlist.1.scale}.

In general, attribute occurrences are named by a grammar symbol, followed by a period, followed optionally by a non-negative decimal integer and another period, followed by the name of an attribute of that symbol. The integer and the second period are needed only when a given grammar symbol appears more than once in the rule, in which case those distinct appearances are numbered from left to right with consecutive increasing integers starting with 0. For a symbol \( X \) with an attribute \( a \), \( X.a \) is a synonym for \( X.0.a \).
A given rule and an attribute occurrence in that rule constitute an attribute occurrence in the grammar.

5 Ox Attribute Definitions

For each rule, the Ox user may provide an attribute reference section, delimited by @\{ and @\}, and optionally containing definitions of attribute occurrences of the given rule. Attribute occurrences may be defined therein in terms of the rule’s other attribute occurrences and C code such as global variables, constants, macros, and function calls.

5.1 Inherited vs. Synthesized Attributes

An attribute occurrence \( o \) of a rule \( R \) is synthesized if and only if

1. \( o \) is on the LHS of \( R \) and the attribute reference section of \( R \) contains a definition of \( o \), or

2. \( o \) is on the RHS of \( R \) and the attribute reference section of \( R \) contains no definition of \( o \).

An attribute occurrence \( o \) of a rule \( R \) is inherited if and only if

1. \( o \) is on the LHS of \( R \) and the attribute reference section of \( R \) contains no definition of \( o \), or

2. \( o \) is on the RHS of \( R \) and the attribute reference section of \( R \) contains a definition of \( o \).

An error message is issued if an attribute is found to have both synthesized and inherited occurrences in the grammar. An attribute is synthesized if and only if it has at least one occurrence, and its every occurrence is synthesized. An attribute is inherited if and only if it has at least one occurrence, and its every occurrence is inherited. It follows from the above that the grammar’s start symbol may have only synthesized attributes. Referring to returned tokens as rules emphasizes the equal status of tokens and non-terminals, inasmuch as each kind of symbol (except the start symbol) may
have both synthesized and inherited attributes. Since each symbol has a distinct name space, same-named attributes of different symbols are different attributes, and may differ as to whether they are inherited or synthesized.

For each parse-tree node except the root node, two rules of the Ox input specification are of particular interest. The home rule is the rule applied at the node, i.e., the rule whose LHS is the label of the given node, and whose RHS symbols are the labels of the children of the node. The parent rule is the rule applied at the node’s parent. The attribute definition of a synthesized attribute instance of a given node is associated with the node’s home rule (i.e., it appears in the attribute reference section for that rule), and definitions of inherited attribute instances are similarly associated with the parent rule.

In a legal input specification, each attribute of a symbol appearing in a rule is either synthesized or inherited, but not both, so the definitions of all attributes “fit together” completely and without contradiction.

5.2 Attribute Reference Sections in the Y-file

The rules section of a Yacc file follows the first % mark [Johnson 75], and contains the productions (rules) of the grammar. As mentioned above, the Ox user may augment each rule by an attribute reference section, each of which is delimited by @{ and @} , and which contains zero or more attribute definitions. When present, the attribute reference section is the last item (other than a terminating semicolon) in a rule. Thus it does not precede any Yacc action or the Yacc reserved word %prec in the rule, and any following identifier must be the LHS of the next rule. Conceptually, an attribute definition has a dependency part and an evaluation part, but syntactically, the parts may be combined or separate. There are three modes of expression of attribute definitions, and different modes may be used within a single attribute reference section. Each attribute definition begins with a definition mode annunciator (@e, @i, or @m), and is terminated by another mode annunciator or by @}.

5.2.1 Explicit Mode

In this, the most powerful and most verbose attribute definition mode, an attribute definition takes the form of @e (mnemonic for explicit) followed by
a *dependency expression* (which expresses the dependency part of the definition) followed by an *evaluation expression* (which expresses the evaluation part). In the following example, the attribute reference section contains three attribute definitions, each expressed in the explicit mode:

```c
num : bitlist DOT bitlist
  @ { @e num.value : bitlist.0.value bitlist.1.value;
      @num.value@ = @bitlist.0.value@ + @bitlist.1.value@;
      @e bitlist.0.scale : ;
      @bitlist.0.scale@ = 0 ;
      @e bitlist.1.scale : bitlist.1.length ;
      @bitlist.1.scale@ = -@bitlist.1.length@;
  @}
```

A dependency expression explicates constraints on the order of execution of evaluation expressions and is a non-empty list of attribute occurrences of the rule, followed by a colon, followed by a possibly empty list of attribute occurrences and a terminating semicolon. The occurrences to the left of the colon are said to *depend upon* (hence are called *dependents of*) those to the right, and are the occurrences *defined* in the given attribute definition. The occurrences to the right are *dependees of* those on the left. An evaluation expression is basically a C code fragment that may contain *attribute references*, each of which is an attribute occurrence enclosed within `@` symbols. Attribute references behave as C variables, and all of the usual C operators, such as those for arithmetical, logical, and pointer operations, may be applied to them, as in a C program. The evaluation expression immediately follows the semicolon of the dependency expression.

The Ox-generated evaluator chooses an evaluation order such that the evaluation expressions for all of the dependees in the definition are executed before those of the dependents. Usually there is a single dependent in a given attribute definition, but in some cases, code may be made more compact by placing more than one attribute occurrence in a dependent list, thereby combining the definitions of those in the list. The evaluation expression is executed on behalf of the dependents *taken as a set*, rather than once for each dependent. This is known as *solving* the attribute instances corresponding to the occurrences in that set.
5.2.2 Implicit Mode

The *implicit mode*, which is the usual mode of expressing attribute definitions, syntactically combines the dependency part with the evaluation part. The following Ox code is equivalent to that of the preceding example.

```ox
num : bitlist DOT bitlist
{ @i @num.value@ = @bitlist.0.value@ + @bitlist.1.value@;
  @i @bitlist.0.scale@ = 0;
  @i @bitlist.1.scale@ = -@bitlist.1.length@;
}@;
```

In this mode, an attribute definition takes the form of @i followed by an evaluation expression. The mode annunciator @i informs Ox that the definition has a single dependent, namely the first attribute occurrence referenced in the evaluation expression. The dependees in the definition consist of all *other* attribute occurrences referenced in the evaluation expression.

5.2.3 Mixed Mode

*Mixed mode* attribute definitions are announced by the reserved word @m, and there follow one or more dependents, a semicolon, and an evaluation expression. The occurrences referenced in the evaluation expression, except those that also appear between @m and the semicolon, are taken to be the dependees in the definition. Thus the dependents are given explicitly and the dependees implicitly. The code in the following example has the same meaning as that in the previous two.

```ox
num : bitlist DOT bitlist
{ @m num.value ;
  @num.value@ = @bitlist.0.value@ + @bitlist.1.value@;
  @i @bitlist.1.scale@ = -@bitlist.1.length@;
  @m bitlist.0.scale ; @bitlist.0.scale@ = 0;
}@;
```

5.3 Attribute Reference Sections in the L-file(s)

Definitions of inherited attributes of tokens are associated with rules appearing in the Y-file, while their synthesized attributes are defined in the L-file(s).
Ox processes the Y-file before processing the L-file(s). If a given attribute occurrence of a token is not defined in the Y-file, then the attribute is taken to be synthesized.

Lexical rules are associated with `return` statements in Lex actions. After the terminating semicolon of each such statement, there may appear a possibly empty `attribute reference section`, delimited by `{` and `}`, in which are defined all of the synthesized attributes of the returned token.

The class of attribute grammars accepted by Ox is restricted only as follows: synthesized attributes of tokens do not have dependees. Attribute definitions in the L-file(s) can thus be written more simply than in the Y-file; each attribute occurrence is defined by referring to it in C code, exactly once in the attribute reference section associated with the `return` statement, as in the following example.

```
[0-9]+    return(CONST); @ { sscanf(yytext,"%d",&CONST.val0); } 
```

Thus mode declarations and dependency expressions are unnecessary in the L-file(s).

A difficulty arises in rules like

```
return(*yytext);
```

and

```
return(cond?TOKEN1:TOKEN2);
```

for which Ox cannot determine at evaluator-generation time which token will be returned. Ox issues a warning in such cases. In the first case, wherein no declared token or character constant is recognized in the `returned` expression, Ox assumes that the token returned has no attributes. In the second case, wherein more than one declared token or character constant is recognized, the node appended to the tree during evaluation is of the type of the declared token or character constant appearing `first` in the expression.

### 5.4 Cycles

It is easy to write an attribute grammar such that some attribute instance of some parse tree has a chain of dependencies that leads back to itself. Such a grammar is called *circular*, and such a chain of dependencies is called
a cycle. For such a tree, there is an attribute instance that the evaluator cannot begin to solve until that instance has already been solved. A cycle also makes it impossible to solve any attribute instance that has a chain of dependencies leading to an instance involved in the cycle. Circularity is usually not intended by the evaluator designer. A general circularity test performed at evaluator-generation time would require exponential running time for some inputs [Jazayeri 75]. Polynomial-time tests for special kinds of non-circularity are known, but the present version of Ox deals with the problem by checking for cycles at evaluation time.

6 Translation into C Code

Ox translates attribute declarations into C structure declarations, with the attribute names appearing as structure members.

The evaluation expression of each attribute definition is copied verbatim into Ox's output, except that attribute references are translated into parenthesized references to C variables.

7 Temporal Behavior of the Ox-generated Evaluator

7.1 Stack Operations

Inasmuch as an ordinary Yacc/Lex recognizer employs an LR parsing algorithm [Aho 86], each input entails a sequence of lookaheads, shifts, and reductions, and a stack of parser states is maintained. From ordinary Yacc/Lex source, Ox generates an evaluator whose yyparse goes through the same sequence of lookaheads, shifts, and reductions as does the yyparse of the ordinary Yacc/Lex recognizer.

The Ox-generated evaluator, in building a parse tree, maintains a stack of subtrees. The operations on the stack of subtrees are synchronized with the operations yyparse performs on its stack of parser states, except that operations involving the “marker nonterminals” (see [Johnson 75]) inserted into the grammar by Yacc are ignored.
The evaluator maintains its stack of subtrees as follows. Lookaheads coincide with calls to \texttt{yylex}. Just before a \texttt{return} is executed in a \texttt{Lex} action, an image of a leaf node is created in the evaluator's \textit{lookahead buffer}, and its synthesized attribute instances are solved and placed in that buffer. At each shift, a leaf node is created from the image in the lookahead buffer, and the subtree consisting of that leaf node is pushed onto the stack. At each reduction, zero or more subtrees are popped from the stack, and their roots become the children of a newly-created node, yielding a new subtree. The root of the new subtree is given a label to indicate the production being applied at the node, and the new subtree is pushed onto the stack. The parse tree is completed upon end of input together with reduction to the start symbol.

\section*{7.2 Placement of Generated Code}

Code for parse-tree management and attribute evaluation is placed in \texttt{Yacc} and \texttt{Lex} actions in \texttt{Ox}'s output. If a given rule in the Y-file has an ordinary \texttt{Yacc} action, the \texttt{Ox}-generated code is placed \textit{after} any programmer-supplied \texttt{C} code contained in the action. If a given rule in the Y-file lacks a \texttt{Yacc} action, an action is created, and the \texttt{Ox}-generated code is placed there. The actions so created are introduced only at the \textit{ends} of rules, so \texttt{Yacc} does not create a marker nonterminal for the action, and the LALR(1) property of the grammar is unaffected.

When an attribute reference section in an L-file contains definitions for more than one attribute occurrence, code for implementing those definitions is executed in the same order in which the definitions appear in that section.

For the attribute occurrences defined in the Y-file, \texttt{Ox} and the \texttt{Ox}-generated evaluator perform analyses to determine when to execute the code segment that evaluates a given attribute. The order of execution of the code segments associated with the definitions in a given attribute reference section is determined by the dependencies of the definitions, and is not necessarily related to the order of appearance of the definitions.

Some attribute occurrences, for example those that have no dependees, are evaluated as part of the \texttt{Yacc} action executed upon reduction by the associated production. Definitions of such occurrences are allowed to contain references to the \texttt{Yacc} pseudovariables \texttt{$\$\$, $1$, $2$}, etc. If \texttt{Ox} determines that a given attribute occurrence cannot be evaluated at reduction time, and the
definition refers to such a pseudovariable, Ox issues an error message.

7.3 Decoration and the Ready Set

The Ox-generated evaluator maintains a set of attribute instances that are ready to be solved, i.e., those whose every dependee has been solved, but which have not themselves been solved. During parsing of the input, it is possible to remove an attribute instance from this ready set, solve it, and then check whether the solving of that instance has caused any of its dependents to be ready to be solved. Instances that are thus made ready are then placed in the ready set. Repeating this process until the ready set is empty is known as decoration. Following a decoration, further parsing of the input may result in creation of parse-tree nodes and insertion of attribute instances into the ready set. Scheduling of decorations is performed automatically by the evaluator. Evaluation of a given syntactically-correct input involves at least one decoration, that performed after the final reduction to the start symbol.

8 Programming Style

Definitions of attribute grammar, (for instance those in [Lorho 88] and [Waite 84]) employ no notion of execution sequence. The usual Ox programming style involves defining synthesized attribute occurrences of tokens in terms of yytext and yyleng and other such data structures of the scanner. Then the attribute definitions of each production are written only in terms of constants and other attribute occurrences of that production. For a given sentence, the synthesized attribute instances of the tokens then completely determine the values of all attribute instances of the parse tree. The attribute instances of the root node are often of particular interest, and their definitions often contain code that copies their values to global C variables, so that they may be used in code executed after the return from yyparse.

Since attribute definitions in Ox code may contain any C code, the Ox programmer may deviate from the safe approach described above by using non-root attribute definitions that read or write global variables. Before attempting the use of side effects, the programmer should be familiar with the material of section 7.
Since the order of evaluation of attributes by the Ox-generated evaluator is not explicit in the Ox input specifications, usually it is not convenient to use attribute definitions for order-sensitive side effects such as code generation.

A common general approach to translation is to build and decorate a parse tree or syntax tree (meanwhile performing some of the checks for semantic errors), and to then make one or more determinate-order tree traversals for final error checks, gathering of compilation statistics, code generation, etc.

Ox has a facility for specification of such traversals, and this is the topic of section 9.

9 Postdecoration Traversals

The idea of decoration was described in section 7.3. Postdecoration refers to any time after the final decoration of the parse tree, which follows parsing of a correct input.

9.1 Example: Infix to Prefix Translation

The problem of parsing infix arithmetic expressions, and their translation to prefix form serves to introduce Ox's postdecoration traversal facility.

The tokens of the example language are determined by the following L-file:

```%
#include "y.tab.h"
#include "ox.out.h"
%

%
[ \n\t\f]*    ;
[0-9]+     return(CONST); @{ sscanf(yytext,"%d",&@CONST.val@); @} \
(          return( '('); \\ .      return( '+'); \\ *      return( '*' );
     .      fprintf(stderr,"illegal character\n");
```
The following Y-file completes the specification of the evaluator.

```yacc
%token CONST
%left '+'
%left '*'

@attributes {int val;} CONST
@traversal @lefttoright @preorder LRpre

{%
#include "ox.out.h"
#include <stdio.h>
%}
```
9 POSTDECORATION TRAVERSALS

```
%%
expr : expr '+' expr  /* rule 2 */
   @{ @LRpre printf(" + "); @}
| expr '*' expr  /* rule 1 */
   @{ @LRpre printf(" * "); @}
| '(' expr ')'  /* rule 3 */
| CONST  /* rule 4 */
   @{ @LRpre printf("%d", @CONST.val()); @}
;

```

```
main()
{ yyparse();
   printf("\n\n");
}
```

The sequence: `@traversal @lefttoright @preorder LRpre` specifies that a left-to-right preorder traversal of the parse tree be performed by the evaluator after the final decoration, and that the traversal be identified as `LRpre`. Note that `LRpre` is programmer-defined, and is not an `Ox` reserved word.

Each attribute reference section in the above Y-file contains a traversal action specifier starting with the traversal mode annunciator `@LRpre`, which is defined in the above-mentioned `@traversal` specification.

When the `LRpre` traversal reaches a node at which rule 1 is applied, an asterisk is printed, then each child of the node, in left-to-right order, is traversed. The behavior of the traversal at a node at which rule 2 is applied is the same, except that a plus sign is printed instead of an asterisk. When `LRpre` reaches a node for rule 3, no traversal action is performed, but the children of the node are traversed recursively as described above for nodes for rules 1 and 2. The `val` attribute of the `CONST` child is printed when a node for rule 4 is reached. When a traversal reaches a terminal node, no action is performed, and the traversal’s recursion degenerates.
9 \textsc{POSTDECORATION TRAVERSALS} \hfill 18

9.2 General Description

9.2.1 Traversal Specifications

The \texttt{Ox} programmer may place in the declarations section (the part before the first `%%' mark) of the Y-file one or more \textit{traversal specifications}. Such a specification consists of the reserved word \texttt{@traversal}, followed by a \textit{traversal specifier sequence} and a non-empty sequence of identifiers, the identifiers being separated by whitespace. A traversal specifier sequence consists of the following (in any order):

- at most one of: \texttt{@postorder}, \texttt{@preorder}
- at most one of: \texttt{@lefttoright}, \texttt{@righttoleft}
- optionally: \texttt{@disable}

If neither \texttt{@postorder} nor \texttt{@preorder} appears in the sequence, the traversal is postorder by default. A left-to-right traversal is specified by default when neither \texttt{@lefttoright} nor \texttt{@righttoleft} appears.

Following the final decoration, the parse tree is traversed once for each traversal specification. The order of performing the traversals corresponds to the order of appearance of the traversal specifications. The \texttt{@disable} reserved word causes the generated evaluator to skip any traversal in whose specification it appears, which may be useful for debugging.

The code fragment:

\begin{verbatim}
@traversal @preorder LRpre
@traversal LRpost
\end{verbatim}

appearing in the declarations section specifies that, after the final decoration, the generated evaluator is to perform a left-to-right preorder traversal named \texttt{LRpre}, followed by a left-to-right postorder traversal named \texttt{LRpost}.

9.2.2 Traversal Action Specifications

In addition to attribute definitions (section 5.2), the attribute reference sections of the Y-file may contain \textit{traversal action specifications}. Each of these consists of a \textit{traversal mode annunciator}, followed by a sequence of \textit{dynamic traversal modifiers} and a \textit{traversal action}. A traversal mode annunciator is \texttt{@} followed immediately by the name of a previously-declared traversal.
Suppose traversal specifications of LRpre and LRpost as above. Then in the code fragment

```c
s : expr
{ @LRpost printf("\n"); /* 1 */
   @LRpost @revorder (1) printf("postfix: "); /* 2 */
   @LRpre @revorder (1) printf("\n"); /* 3 */
   @LRpre printf("prefix: "); /* 4 */
}
```

the attribute reference section has four traversal action specifications and no attribute definitions. Each specification is announced by either @LRpre or @LRpost. Each of the printf statements constitutes a traversal action. The form of a traversal action is that of a C code fragment, except that it may contain references to the attribute occurrences of the associated rule.

The second and third specifications each have @revorder (1) as a dynamic traversal modifier. A dynamic traversal modifier is either @revorder or @revdirection, followed by a parenthesized expression that conforms to C syntax, except that it may refer to the rule’s attribute occurrences. @revorder and @revdirection may each occur at most once in a given traversal action specification. If @revdirection appears in two traversal action specifications within a given attribute reference section, the two specifications must have different annunciators. Dynamic traversal modifiers are used to override the traversal specifications of a given traversal when it reaches a given kind of node. The modifier @revorder expr means roughly “reverse order if expr”. When the LRpre traversal reaches a node at which the rule s : expr is applied, the expression (1) is evaluated, and because it is nonzero, the third traversal action, which prints a line feed, is executed as if LRpre were a postorder traversal, i.e., after the recursive traversal of the subtree rooted at the node’s sole child. The execution of the fourth traversal action, printf("prefix: "); is not affected by any dynamic traversal modifier, and occurs according to LRpre’s (static) specification, i.e. before the traversal of the child subtree.

When the LRpost traversal reaches a node at which s : expr is applied, the second traversal action is executed, the traversal proceeds to the child subtree, then the first traversal action is executed.
9 POSTDECORATION TRAVERSALS

9.2.3 Traversal Semantics

The behavior of postdecoration traversals was illustrated in the preceding examples. In view of those examples, the following C-like pseudocode holds no surprises, but describes such behavior generally and precisely. The traversals are carried out by a single call of doTraversals (below) after the final decoration.

```c
enum orderType {PREORDER, POSTORDER};
enum directionType {LEFTTORIGHT, RIGHTTOLEFT};

enum orderType staticOrder(traversal T){
    if (@preorder appears in the traversal specification of T)
        return PREORDER;
    return POSTORDER;
}

directionType staticDirection(traversal T){
    if (@righttoleft appears in the traversal specification of T)
        return RIGHTTOLEFT;
    return LEFTTORIGHT;
}

int isDisabled(traversal T){
    if (@disable appears in the traversal specification of T)
        return 1;
    return 0;
}
```
void pdTrav(parse_tree_node N, traversal T)
{
    grammar_rule R; /* the rule applied at N */
    enum orderType order[Z]; /* Z >= # of traversal action specs */
        for T in R */

    enum directionType direction;
    int i,j,k;

    R = the grammar rule applied at N;
    let the traversal actions for T in the attribute definition section
        of R be numbered from 0 to k-1;
    for (i=0; i<k; i++)
    {
        if (the ith traversal action specifier has no @revorder)
            order[i] = staticOrder(T);
        else if ((the expression associated with @revorder) == 0)
            order[i] = staticOrder(T);
        else if (staticOrder(T) == POSTORDER)
            order[i] = PREORDER;
        else
            order[i] = POSTORDER;
        if (the ith traversal action specifier has no @revdirection)
            direction = staticDirection(T);
        else if ((the expression associated with @revdirection) == 0)
            direction = staticDirection(T);
        else if (staticDirection(T) == LEFTTORIGHT)
            direction = RIGHTTOLEFT;
        else
            direction = LEFTTORIGHT;
    }
    for (i=0; i<k; i++)
    {
        if (order[i] = PREORDER)
            execute the ith traversal action;
        number the children of N from left to right
            with integers from 0 to j-1;
        if (direction == LEFTTORIGHT)
            for (i=0; i<j; i++) pdTrav(the ith child of N, T);
        else
            for (i=j-1; i>0; i--) pdTrav(the ith child of N, T);
        for (i=0; i<k; i++)
        {
            if (order[i] = POSTORDER)
                execute the ith traversal action;
        }
    }
`doTraversals()`

```c
void doTraversals()
{
    int i, k;
    parse_tree_node r;

    r = the root of the parse tree;
    k = the number of traversals;
    number the traversals from 0 to k-1, according to
    the order of appearance of their specifications;
    for (i=0; i<k; i++)
        if (!isDisabled(the ith traversal))
            pdTrav(r, the ith traversal);
}
```

## 10 Macros

Ox's input specification may be such that the same or similar text appears in more than one place in attribute reference sections. Ox has a macro substitution feature that can be used to decrease verbosity in such cases.

### 10.1 Macro Definitions

Ox macros are defined in the declaration section of the Y-file. Such a definition consists of the `@macro` reserved word, an identifier (the name of the macro), a left parenthesis, a parameter list, a right parenthesis, the body of the macro, and the `@end` reserved word. The parameter list is a possibly empty sequence of identifiers, each (including the last, if the list is nonempty) followed by a comma. Each identifier is a sequence of letters and digits, beginning with a letter. The body of the macro is a segment of arbitrary text, terminated by the first occurrence of `@end`, with the following exceptions: When inside a comment or a string, or when preceded immediately by the backslash escape character, an occurrence of `@end` is considered part of the macro body (hence does not terminate the macro). Such a backslash character is deleted from the macro body.
10.2 Macro Uses

Ox macros are used only in attribute reference sections and in other Ox macros. Substitution occurs where a macro use is encountered outside of a string, comment, or attribute name.

A macro use consists of the name of a previously-defined macro, and an argument list in parentheses. The argument list is a possibly empty sequence of text fragments, each (including the last) such fragment terminated by a comma. In expanding a macro use, each text fragment is substituted for each occurrence in the macro body of the corresponding parameter in the macro definition. If commas, parentheses, or backslashes are to appear in a text fragment, they must be preceded by backslash escape characters, which are removed during substitution.

It is not necessary that the definition of a macro precede that of another macro in which it is used, as no macro substitution occurs until Ox processes the attribute reference sections.

10.3 Example

The following excerpts from a Y-file illustrate the use of Ox macros.

```
@end

@macro exprdefs(op,)
  @i @expr.1.env@ = @expr.env@;
  @i @expr.2.env@ = @expr.env@;
  @i @expr.type@ = typeResolve(@expr.1.type@, @expr.2.type@,);
  @i @expr.value@ = exprEval(op, @expr.type@, @expr.1.type@, @expr.2.type@, 
                             @expr.1.value@, @expr.2.value@
                           )
@end

@macro typeResolve(type1, type2,)
  ((type1 == type2)?type1:FLOATTYPE)
@end
```

```
11 Command Sequences for Evaluator Development

11.1 Conventions of Naming Ox Output Files

Ox translates the Y-file into a file destined for processing by Yacc, given the name ox.out.y. The I-files are translated into files destined for Lex. If there is exactly one I-file, its corresponding output file is named ox.out.1. If there is more than one I-file, the corresponding outputs are named ox.out.0.1, ox.out.1.1, ox.out.2.1, etc.

11.2 Review: Combining the Outputs of Yacc and Lex

In developing an ordinary (i.e., non-Ox) Yacc/Lex evaluator, y.tab.c and lex.yy.c can be compiled immediately into an executable file by placing the line

```
#include "lex.yy.c"
```

in a C-code section of the Yacc input specification [Leks 75]. Alternatively, Yacc can be instructed (by using the -d command-line option) to produce a separate file y.tab.h that contains declarations needed by
both \texttt{y.tab.c} and \texttt{lex.yy.c}. The two files may then be compiled separately if the line

\begin{verbatim}
#include "y.tab.h"
\end{verbatim}

is placed in C-code sections of both the \texttt{Yacc} and the \texttt{Lex} input specifications. The two resulting object files can then be linked to produce an executable file.

### 11.3 Combined Use of \texttt{Ox}, \texttt{Yacc}, and \texttt{Lex}

There are certain declarations that must be visible from all of the files produced by \texttt{Ox}. By default, \texttt{Ox} produces files suitable for separate compilation, inasmuch as the \texttt{Yacc}-destined file and the \texttt{Lex}-destined file(s) each contain the common declarations. \texttt{Ox} also supports the one-step development approach described above. By placing \texttt{-h} on \texttt{Ox}'s command line, the designer calls for generation of a file \texttt{ox.out.h} containing the common declarations, which are then absent from \texttt{Ox}'s other output files. In this case, the line

\begin{verbatim}
#include "ox.out.h"
\end{verbatim}

is placed in the Y-file.

### 11.4 Typical Command Sequences

The following sequence of shell commands is an example of the separate compilation approach described. In this example, \texttt{Ox} translates the Y-file \texttt{ev.Y} into \texttt{ox.out.y} and the L-file \texttt{ev.L} into \texttt{ox.out.l}. The last command of the sequence links the two object files, yielding the executable file \texttt{ev}.

\begin{verbatim}
ox ev.Y ev.L
yacc -d ox.out.y
lex ox.out.l
cc -c y.tab.c
cc -c lex.yy.c
cc -o ev.y.tab.o lex.yy.o -ll -ly
\end{verbatim}

The following command sequence does a one-step compilation.

\begin{verbatim}
ox -h ev.Y ev.L
yacc ox.out.y
lex ox.out.l
cc y.tab.c -ll -ly
\end{verbatim}
12 Other Points and Features

12.1 Error Recovery

Yacc has provisions for building parsers that attempt to recover from syntax errors, and the designer can use the words error, yyerror, and yyclearin to implement such error recovery [Johnson 75]. When a parser that employs such techniques detects a syntax error, it may attempt to recover by popping items from its stack or by discarding tokens. During normal parsing, the Ox-generated evaluator synchronizes its stack operations with those of the Yacc-generated parser (see section 7). When the parser is built using error, yyerror, or yyclearin, and a syntax error occurs, this synchronization is lost. It is possible for the evaluator to corrupt its stack and go out of control in such cases. Ox provides the function yyyerror to prevent such chaos. The parser calls yyerror upon any syntax error, and the designer should write yyerror such that yyyerror is executed at least once each time yyerror is called. Any syntax error will then cancel parse-tree construction and attribute evaluation, and it is ensured that the Yacc-generated code can continue safely. Use of yyyerror is unnecessary but harmless if the Y-file makes no use of the words error, yyerror, and yyclearin.

12.2 Memory Alignment

Many computing systems have hardware-related constraints on the addresses used for memory accesses. For example, for a certain type it may be required that the first byte of storage for each variable of that type reside at an even-numbered address. Then an instruction to access a variable of that type at an odd-numbered address results in a run-time error. When Ox is given the -aN command-line option, it produces an evaluator that aligns all C structures on addresses divisible by the integer N. The default value for this alignment constant is 4, which is adequate for nearly all current computers.

12.3 Stripping Ox Constructs

Occasionally, the designer may wish copies of the Y-file and L-file(s) free of Ox-specific constructs. Suppose, for instance, that changes to the underlying grammar are under consideration, and that it is desired to test whether the
new grammar has parsing conflicts. To satisfy Ox semantics might require writing attribute definitions for any new rules. Ox's output on ox.out.y could then be submitted to Yacc to test for parsing conflicts.

To avoid the above-mentioned writing of attribute definitions, the designer can use Ox's -s command-line option, which filters all Ox-specific constructs from the inputs and yields files acceptable to Yacc and Lex. The original copies of the Y-file and L-file(s) are unchanged, but Ox's outputs on ox.out.* contain neither Ox constructs nor the usual Ox-generated parse-tree management code.

12.4 Preventing Execution of Attribute Definition Code

Faulty user-written code in attribute reference sections may cause abnormal termination of the Ox-generated evaluator. For instance, dereferencing a stray pointer may corrupt the evaluator's data structures and cause it to falsely report a cycle during attribute evaluation. The -n command-line option is a debugging feature that can be used to isolate the effects of anomalous attribute definition code. When Ox is used with this option, the generated evaluator uses the ready set as usual to determine an evaluation order for attribute instances, and still checks for cycles. Each time it is ready to solve an instance, however, it stops short of executing the code for the definition of that instance. When -n is used, the designer should take special notice of the effects upon other translation phases of such suppression of semantic analysis.

12.5 Control of Storage Allocation in the Generated Evaluator

When initializing itself, the Ox-generated evaluator allocates memory for its various data structures. When evaluating a large input, it may happen that the space allocated for a given data structure is inadequate. In such a case, the evaluator issues an error message indicating which data structure was exceeded and suggesting an appropriately larger size. The sizes of these data structures may be determined by the default values built into Ox, or by the evaluator designer's use of the -Ya N option on the Ox command line, where a is an alphabetic character that specifies the data structure to be sized, and N is an integer that determines the size of data structure a.
The evaluator designer can easily build an evaluator that accepts the same \(-\text{-YY}\) command-line options accepted by \texttt{Ox}. By specifying the \(-\text{-YY}\) option on \texttt{Ox}'s command line, \texttt{Ox} is instructed to declare in the generated evaluator a function \texttt{yyyCheckForResizes} that can read \texttt{main}'s arguments (i.e., the command-line options passed to the generated evaluator) and adjust sizes accordingly. When using the \(-\text{-YY}\) option, the designer should arrange the evaluator’s \texttt{main} program according to the following form:

```c
void yyyCheckForResizes();

main(argc, argv)
    int argc;
    char *argv[];
{

    /* This is executed before calling yyparse */
    yyyCheckForResizes(argc, argv);

    yyparse();

}
/* main */
```

### 12.6 Parse Tree Statistics

Placing \(-\text{-u}\) on \texttt{Ox}'s command line causes generation of an evaluator that prints, for each input, statistics regarding the parse tree built for the input. These include numbers of:

- terminal nodes and their attribute instances,
- nonterminal nodes and their attribute instances,

and other statistics.
12.7 Adjusting the Sizes of Ox’s Data Structures

Ox itself calls system memory allocation routines to obtain storage for its internal data structures. The default sizes of these data structures are quite generous, and exceeding them would be somewhat unusual. In case any of these is exceeded, Ox prints an error message indicating the use of a command-line option of the form \(-X aN\) to make \(N\) the size of data structure \(a\).

13 Example: An Integer Calculator

This section has Ox code for an evaluator of simple expressions involving multiplication and addition. Since the grammar has only synthesized attributes, the Ox implementation offers little advantage over one that uses only Yacc and Lex; it is presented as a very easy example of Ox usage.

The L-file specifies that the tokens are digit strings, parentheses, ‘*’, and ‘+’:

```c
%{
/* expr.L: L-file for a simple expression language */
#include "y.tab.h"
#include "ox.out.h"
%
%
[ \m\t\f]*
[0-9]*
return(CONST); @{
    sscanf(yytext,"%d",&@CONST.val@)); @}
\( return('('*); \\
\) return(')'); \\
\+ return('+'); \\
\* return('*'); 
%
```
The grammar is disambiguated by use of Yacc's \%left reserved word. Each parse-tree node labeled by s, e, or CONST has an integer attribute instance named val. Use of the global variable sVal obviates postdecoration traversal.

/* expr.Y: Y-file for a simple expression language */
%left '+'
%left '*'
%token CONST

@attributes {long val;} s e CONST

{%
#include "ox.out.h"
long sVal;
%

%%
s : e
{ i sVal = s.val = e.val; 0}
;

e : e '+' e
{ i e.0.val = e.1.val + e.2.val; 0}
;

e : e '*' e
{ i e.0.val = e.1.val * e.2.val; 0}
e : '(' e ')' e
{ i e.val = e.1.val; 0}
| CONST
{ i e.val = eCONST.val; 0}
;

%%

main()
{yyparse();
 printf("%d\n",sVal);
}
The following command sequence is used to build an executable file `calc` from the above specifications:

```
    ox  -h expr.Y expr.L
    yacc -d ox.out.y
    lex ox.out.l
    cc  -c y.tab.c
    cc  -c lex.yy.c
    cc  -o calc y.tab.o lex.yy.o -ly -ll
```

14 Example: A Binary Number Translator

This illustrates the use of Ox to build an evaluator based on an example attribute grammar that appears in the seminal paper on the subject [Knuth 68]. The input (after removal of whitespace) is either a nonempty string of binary digits or two such strings separated by a period. This input is interpreted as a binary representation of a floating point number, which is then printed on the standard output in its base-ten form.

Following is the text of the L-file:

```
{%
#include "y.tab.h"
%
%
[0]            return ZERO;
[1]            return ONE;
\.
[\n\t\v ]       ;
.             {fprintf(stderr,"illegal character\n");
              exit();
}
Here is the text of the Y-file:

```yacc
%token ZERO ONE DOT

@attributes {float value; int scale;}       bit
@attributes {float value; int scale,length;} bitlist
@attributes {float value;}                num

%start num

{%
#include <stdio.h>
float numValue;
%
```


%%% 

bit : ZERO
    @{ @i @bit.value@ = 0;
        /* value is synthesized for bit. */
        /* scale is inherited for bit. */
    }

bit : ONE
    @{ @i @bit.value@ = twoToThe(@bit.scale@);
        @}

bitlist : bit
    @{ @i @bitlist.value@ = @bit.value@;
        @i @bit.scale@ = @bitlist.scale@;
        @i @bitlist.length@ = 1;
        /* value and length are synthesized for bitlist. */
        /* scale is inherited for bitlist. */
    }

    | bitlist bit
    @{ @i @bitlist.0.value@ = @bitlist.1.value@ + @bit.value@;
        @i @bit.listscale@ = @bitlist.0.scale@;
        @i @bitlist.1.scale@ = @bitlist.0.scale@ + 1;
        @i @bitlist.0.length@ = @bitlist.1.length@ + 1;
        @}

num : bitlist
    @{ @i numValue = @num.value@ = @bitlist.0.value@;
        @i @bitlist.scale@ = 0;
        /* value is synthesized for num. */
    }

    | bitlist DOT bitlist
    @{ @i numValue = @num.value@ = @bitlist.0.value@ + @bitlist.1.value@;
        @i @bitlist.0.scale@ = 0;
        @i @bitlist.1.scale@ = - @bitlist.1.length@;
        @}

%%%
main()
    {if (!yyparse())
        printf("\%30.15f\n", numValue);
    }

float twoToThe(in)    /* returns 2 raised to the power in */
    int in;
    {if (in < 0) return (1.0 / twoToThe(-in));
        if (in == 0) return 1.0;
            else return (2.0 * twoToThe(in - 1));
    }

Job development of the above evaluator follows the separate compilation approach described in section 11.

Removing the Ox-specific constructs and the printf statement from the above source yields a pair of files that constitute a semantics-free recognizer of binary numbers.
15 Example: Translation to Postfix and Prefix

In this example, the generated evaluator is to perform two postdecoration traversals, one for printing the prefix form of a given infix expression, and one for printing the postfix form. The tokens of the language are specified as follows:

```c
#include "y.tab.h"
#include "ox.out.h"

char *lexeme()
{
    char *dum;
    dum = (char *)malloc(yylen+1);
    strcpy(dum,yytext);
    return dum;
}

%%% 

[ \n\t\f]* ;
[0-9]+\.?[0-9]* return(CONST); @{ @CONST.lexeme0 = lexeme(); @}
[A-Za-z_][A-Za-z_0-9]* return(ID); @{ @ID.lexeme0 = lexeme(); @}
/\( return('(');
\) return(')');
\* return('*');
\+ return('+');
- return('-');
%%

```
The first traversal performed is named \texttt{LRpre}, and the second is named \texttt{LRpost}. By default, both are left-to-right traversals. \texttt{LRpost} is a postorder traversal by default. \texttt{LRpre} is specified as a preorder traversal.

/* Y-file for translation of infix expressions to prefix and postfix */
\%token ID CONST
\%start s
\%left '+' '-'
\%left '*' '/'

@attributes {char *lexeme;} ID CONST
@traversal @preorder LRpre
@traversal LRpost

%
#include "ox.out.h"
#include <stdio.h>
%}
```c
/*
expr : expr * expr
    { @LRpost printf(" * ");
      @LRpre printf(" * ");
    }
| expr + expr
    { @LRpre printf(" + ");
      @LRpost printf(" + ");
    }
| expr / expr
    { @LRpre printf(" / ");
      @LRpost printf(" / ");
    }
| expr - expr
    { @LRpre printf(" - ");
      @LRpost printf(" - ");
    }
| '(' expr ')' |
    ID
    { @LRpost printf(" \%s ", @ID.lexeme@);
      @LRpre printf(" \%s ", @ID.lexeme@);
    }
| CONST
    { @LRpost printf(" \%s ", @CONST.lexeme@);
      @LRpre printf(" \%s ", @CONST.lexeme@);
    }
*/

main()
{ yyparse();
 }
```
References


IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY
DEPARTMENT OF COMPUTER SCIENCE

Tech Report: TR 92-30
Submission Date: December, 1992