

OC.doc

COLLABORATORS

	<i>TITLE :</i> OC.doc	
<i>ACTION</i>	<i>NAME</i>	<i>DATE</i>
WRITTEN BY		January 17, 2023
<i>SIGNATURE</i>		

REVISION HISTORY

NUMBER	DATE	DESCRIPTION	NAME

Contents

1	OC.doc	1
1.1	OC.doc	1
1.2	What is OC?	2
1.3	Distribution and Copyright	2
1.4	System requirements	3
1.5	Running OC from the CLI	3
1.6	Running OC from the Workbench	5
1.7	Running OC from the FPE utility	5
1.8	Language extensions supported by the compiler	5
1.9	Escaped characters	6
1.10	Pointer variants	6
1.11	Assignable procedures	7
1.12	Amiga library functions	8
1.13	The pseudo-module SYSTEM	9
1.14	Data types	10
1.15	Memory management	11
1.16	Memory access	11
1.17	Logical operations	11
1.18	Inline machine code	12
1.19	Miscellaneous	12
1.20	Module SYSTEM Reference	12
1.21	Controlling the compiler with switches	14
1.22	Copying open arrays	15
1.23	Index checking	15
1.24	Addressing global variables	15
1.25	Checking memory allocations	16
1.26	Portable code	16
1.27	Range checking	16
1.28	Type checking	17
1.29	Overflow checking	17

1.30	Zeroing variables	17
1.31	Using the garbage collector	17
1.32	Handling run-time errors	19
1.33	Error reports produced by the compiler	19
1.34	Implementation of basic types	19
1.35	Limits built in to the compiler	20
1.36	Who is responsible for THIS?	21
1.37	Reporting bugs and suggestions	21
1.38	Who did what and why	22
1.39	Release history	22

Chapter 1

OC.doc

1.1 OC.doc

\$RCSfile: OC.doc \$

Description: Documentation for the Oberon-A compiler

Created by: fjc (Frank Copeland)

\$Revision: 3.1 \$

\$Author: fjc \$

\$Date: 1994/05/19 23:37:30 \$

Copyright © 1994, Frank Copeland.

Description
What is OC?

Distribution
Copyright and distribution

Requirements
What do I need to run OC

Running OC...

From the CLI
Running OC from the CLI

From the Workbench
Running OC from the Workbench

From FPE
Running OC from the FPE utility

Oberon-2 The programming language Oberon-2

Extensions
Language extensions supported by the compiler

Module SYSTEM

	The pseudo-module SYSTEM
	Compiler switches
	Controlling the compiler
	Garbage collection
	Using the garbage collector
	Run time errors
	Handling run-time errors
	Error reports
	Error reports from the compiler
	Basic types
	Implementation of basic types
	Compiler limits
	Limits built in to the compiler
	The Author
	Contacting the author
	Bugs & Suggestions
	Reporting bugs and suggestions
	Acknowledgements
	Who did what and why
Changes	Changes since the last release
To Do	Bugs to fix and improvements to make
	Release history
	The history of OC

1.2 What is OC?

OC is a (fairly) fast single pass compiler that directly generates MC68000 machine code. The object files it produces are in standard AmigaDOS format and are linkable with BLink.

The compiler translates source code written in the Oberon-2 language described in the Oberon-2 Report by Niklaus Wirth and Hanspeter Mössenböck. It also supports a number of compiler options and language extensions that allow direct access to the Amiga operating system without messy assembler "glue code".

1.3 Distribution and Copyright

OC is part of Oberon-A and is:

Copyright © 1993-1994, Frank Copeland

Parts of OC are based on source code developed at ETH Zuerich. Permission to use, copy, modify and distribute this software is granted by ETH (see the file ETH-Copyright.txt).

See Oberon-A.doc for its conditions of use and distribution.

1.4 System requirements

To use OC requires an Amiga personal computer with at least 1 MB ↔ of RAM, running Kickstart version 1.3 or greater. Depending on the module being compiled, up to 500K of free RAM must be available to run the program. A conscious effort has been made to make sure that OC will run under Kickstart 1.3, but it was developed under Kickstart 2.04 and it was not possible to test this. It is likely that at some future date OC will require Kickstart 2.04 or greater. If this is a problem for you, please contact the author

A linker is needed to create executable programs from the object files generated by the compiler. A freely-distributable linker, BLink, is included in the Oberon-A archive. The ALink linker, available from Commodore, should also work (it has not been tested). Other third-party linkers such as PhxLink may also be suitable.

1.5 Running OC from the CLI

Usage: OC {option} {<filename>}

Options: NS | NEWSYMFIL
DEBUG
VERBOSE
BATCH
{SYM | SYMBOLS <directory>}
DST | DESTINATIONS <directory>

Purpose: To translate an Oberon-2 source text into MC68000 machine code.

Path: Oberon-A/C/OC

OC can be operated in three modes: interactive, command line and batch.

If no filename is given in the arguments passed to OC, it will enter interactive mode and repeatedly prompt the user for the name of a file to be compiled. It will exit when the user presses <enter> in response to the prompt. If a name is entered, it will attempt to compile that file.

If the BATCH argument is passed to OC and one or more filenames have been specified it will enter batch mode. In batch mode OC will attempt to open all the files passed as arguments and interpret their contents as the names of files to be compiled.

If one or more filenames are given and the BATCH keyword is omitted, OC will attempt to compile the files named in the command line.

OC skips anything in a source file before the first "MODULE" symbol. If there isn't one, it will scan the whole file before reporting an error. This feature allows the programmer to include a header in the file which may be meaningful to another translator. For instance, the file might start with a sequence of commands that the AmigaDOS Execute command can interpret as commands to compile and link the module contained in the file.

If any errors are detected, their location and description are output in the file "<module>.err" in the current directory. If there are no errors, an object file (containing machine code, data and relocation information) is output in the current directory, with a name consisting of the name of the module specified in the text plus a ".Obj" extension.

If the compiler cannot find a symbol file for the module, it will output one in the current directory. A symbol file contains information about the constants, types, variables and procedures exported by a module and is used by the compiler if the module is imported by another module. If the NEWSYMFIL option is specified and the module's definition has been changed, the compiler will replace the existing symbol file. If the module's definition has changed and NEWSYMFIL is NOT specified, an error (obsolete symbol file) is reported. The symbol file's name consists of the module name plus a ".Sym" extension.

When searching for symbol files, the compiler will look in the current directory first. If any SYMBOL parameters have been specified, it will then search those directories in the order they have been given. If unsuccessful, it will search "OLIB:". If a DESTINATION parameter is specified, the compiler will output any object and symbol files it generates in the specified directory instead of in the current directory. With both SYMBOL and DESTINATION parameters, the path must end in a "/" or ":" character.

If the DEBUG parameter is specified, the compiler will generate symbol hunks containing the names of all procedures and other objects in the module. These hunks are for the benefit of third-party debuggers. To exclude them from the final program, use the NODEBUG option with BLink or the equivalent if you use another linker.

The VERBOSE and TRACE arguments are for debugging and are undocumented.

The CLI stack should be set to at least 10000 bytes. See the Stack command in the AmigaDOS manual.

Typing in the full command line can become tedious. It is suggested that you adopt a consistent strategy for storing the source, symbol and object files of a project. The author keeps each project in a separate

directory and creates a sub-directory called "Code" to hold the symbol and object files. It is suggested that all library modules' symbol and object files be kept in the "OLIB:" directory, which the compiler automatically searches. A Shell alias can then be created to simplify calling the compiler:

```
alias OComp OC SYM Code/ DST Code/ DEBUG []
```

A module can then be compiled by typing:

```
OComp <module>.mod
```

Other aliases can be created for compiling library modules and doing batch compiles. See the file Oberon-A:S/Oberon-Startup for some suggested aliases.

Examples:

```
OC DST OLIB: DEBUG Intuition.mod  
OC NS SYM Code/ DST Code/ OCE.mod
```

1.6 Running OC from the Workbench

The compiler cannot be run from the Workbench. Sorry. This will come in a future version.

1.7 Running OC from the FPE utility

A tool button in the FPE window can be configured to run the compiler (see FPE.doc). In the button editor, set the Command field to the full path name of the OC program. Set the Arguments field to "!F" plus any options that are desired. Specify a console window as the Console field. Put at least 10000 in the stack field.

For example:

```
Command="DH1:Oberon-A/OC"  
Arguments="!F SYM OLIB: SYM Code/ DST Code/"  
Console="CON:0/11/540/189/Compiling.../CLOSE/WAIT"  
Stack=10000
```

To compile a source file:

1. select the module in the Module gadget.
2. select the file extension from the Files gadgets.
3. click on the tool button the compiler is bound to.
4. sit back and relax for a few seconds.

1.8 Language extensions supported by the compiler

There are two justifications for a compiler allowing deviations from a computer language's formal definition. One is to "improve" the language; the other is to provide machine-dependant facilities. With one exception, all the language extensions supported by this compiler are machine-dependant facilities. In order to use any of these extensions, the global compiler switch \$P must be set to FALSE (see

Compiler Switches
).

Escaped characters

Pointer variants

Assignable procedures

Amiga library functions

1.9 Escaped characters

The Oberon-A compiler allows the use of escaped characters in character and string constants. An escaped character consists of a "\" character followed by one or more characters. The "\" character indicates to the compiler that the following character(s) has special meaning. The meanings are:

```
\0  : insert a nul (NUL, 0X) character.
\b  : insert a backspace (BS, 08X) character.
\t  : insert a tab (HT, 09X) character.
\n  : insert a newline (LF, 0AX) character.
\v  : insert a vertical tab (VT, 0BX) character.
\f  : insert a form-feed (FF, 0CX) character.
\r  : insert a carriage return (CR, 0DX) character.
\xnn : insert the character with ASCII value nn hex.
```

For any other combination, the compiler ignores the "\" character and inserts the following character. So, to insert a "\" character, use the sequence "\\". The most common use for this mechanism is to insert formatting characters in strings to be output to the console, making the task of console IO simpler.

1.10 Pointer variants

Oberon defines only one kind of pointer; this compiler provides two variants. Pointer variants are indicated by alternatives to the normal POINTER keyword. A normal, non-variant pointer is referred to here as an Oberon pointer.

The first pointer variant is known as a CPointer. It is declared with the CPOINTER keyword. It corresponds to the notion of a pointer used by the C language. Such pointers are endemic in the Amiga operating system; this is not surprising since it is largely written in that language. They differ from normal Oberon pointers in two ways: they may have ANY type as a base type; and there is no type tag associated with the object they point to. This last feature means that they cannot be used in type guard statements. Otherwise, they may be used in the same way as any normal Oberon pointer. However, it is recommended that they only be used when declaring objects required by a module providing an interface to the Amiga operating system.

The second variant is known as a BPointer. It is declared with the BPOINTER keyword. It represents a BCPL pointer, a curious object that is found in many data structures used by AmigaDOS. It is a longword pointer; that is, a byte address divided by 4. It is treated by the compiler in much the same way as a CPointer. The compiler automatically generates the code needed to convert between longword and byte addresses when such a pointer is dereferenced.

Examples:

```
TYPE
  Object = POINTER TO ObjDesc;          (* An Oberon pointer *)
  STRPTR = CPOINTER TO ARRAY 32767 OF CHAR; (* A CPointer *)
  FileHandlePtr = BPOINTER TO FileHandleRec; (* A BPointer *)
```

The standard procedure NEW may be used to allocate memory for CPointer and BPointer variables, as well as the SYSTEM.NEW procedure. Both will handle the special requirements of BPointers transparently. SYSTEM.DISPOSE can be used to free the memory once it is no longer required.

You are also free to use the Amiga memory allocation functions such as AllocMem() and FreeMem() with CPointer and BPointer variables. Note that if you do this with BPointers, you will have to handle the necessary shifting yourself.

DO NOT USE AMIGA MEMORY ALLOCATION FUNCTIONS WITH OBERON POINTERS. The compiler makes assumptions about the structure of the memory allocated to an Oberon pointer that are only valid if it is allocated with NEW. The compiler will not allow you to use SYSTEM.NEW with Oberon pointers where this is inappropriate.

1.11 Assignable procedures

Procedures that are to be assigned to procedure variables must be marked with a "*" character, unless they are marked as exported. This tells the compiler to treat them as if they were exported, without making them visible outside the module. This mainly involves generating code to save, set and restore register A4 as the global variable base pointer. The mark character must be placed immediately after the PROCEDURE keyword. For example:

```
PROCEDURE * Assignable;
  ^
  Mark character
```

1.12 Amiga library functions

Amiga system software is accessed through shared code libraries. An Amiga shared library consists of a block of variables and a table of jump instructions. There is one of these jump instructions, known as a function vector, for each function provided by the library. Each vector is accessed by a negative offset (known as the function vector offset) from the base of the library's variables. A library function is called by placing the address of the library variables in register A6 and coding a "JSR offset(A6)" instruction, where "offset" is the vector offset of the desired function. Parameters are placed in specific registers before the function call and results are also returned in registers. See the Amiga ROM Kernel Manual for an in-depth discussion of this process.

The simplest method for a compiler to interface with Amiga library calls is to require that the programmer declare a normal procedure and use assembly language stubs or facilities such as SYSTEM.SETREG to set up the parameters and make the call. This is an inefficient and error-prone system and most recent compilers, including Oberon-A, provide a means for describing library calls in such a way that the compiler can generate the call directly.

The syntax for declaring Amiga library calls used by OC is similar to the syntax for type-bound procedures in Oberon-2, which Amiga library calls closely resemble.

The formal syntax is:

```
$ LibCallDeclaration = LibCallHeading ";"
$ LibCallHeading = PROCEDURE "-" Receiver identdef
  [LibCallParameters] ["-"] integer.
```

The integer value is the library function vector offset. It may be negative; if not, the compiler will negate it anyway.

```
$ Receiver = "(" ident ":" ident ")".
```

The first ident is a dummy. The second must be a type identifier that has been declared earlier in the same module. The type must be a CPOINTER TO RECORD (a CPointer to a record, see Pointer variants) and should be an extension of the Exec.LibraryPtr type (Amiga Resources are extensions of the Exec.NodePtr type).

```
$ LibCallParameters = "(" [LCPSection {";" LCPSection}] ")"
  [":" qualident].
$ LCPSection = [VAR] ident RegSpec {"," ident RegSpec } ":"
  FormalType.
$ RegSpec = "{" integer "}"
```

The integer in a RegSpec must be in the range 0 .. 15 and it

represents a CPU register number. The data registers D0 .. D7 are numbered 0 .. 7; the address registers A0 .. A7 are numbered 8 .. 15. It is used to indicate which register the corresponding parameter is to be passed in.

A less formal example is:

```
PROCEDURE- (base : ExecBasePtr) OpenLibrary*
  ( libName {9} : ARRAY OF CHAR;
    version {0} : Types.ULONG )
  : Types.APTR;
- 552;
```

This defines the Amiga Exec library function `OpenLibrary`. It indicates that it is bound to the `ExecBasePtr` type. It is marked for export. It has two parameters: `libName` is an `ARRAY OF CHAR` whose address is to be passed in register A1; `version` is a `ULONG` (effectively a `LONGINT`) to be passed by value in register D0. The function returns an `APTR` value. Its jump vector can be found 552 bytes before the library base address.

A library procedure is bound to a library base type (which is specified in parentheses before the procedure name). This means that to call it you must first declare a variable of the library base's type and initialise the variable with the address of the base of the corresponding shared code library (use the `Exec OpenLibrary` function for this). The procedure is then called as if it were a type-bound procedure bound to the library base type.

Example:

Assuming that it has been declared in module `Exec`, along with a variable called `Base` of type `ExecBasePtr`, a call of `OpenLibrary ()` might look like this:

```
DiskFontBase := Exec.Base.OpenLibrary ("diskfont.library", 33);
```

Library procedures are NOT inherited by extensions of the type they are bound to.

1.13 The pseudo-module SYSTEM

Every Oberon implementation includes a pseudo-module called `SYSTEM`, defined internally in the compiler. Its purpose is to provide machine-dependant and low-level facilities that cannot otherwise be expressed in the Oberon language. The `SYSTEM` module provided with Oberon-A is based on the module defined for the Ceres compiler but contains several differences.

Data types
 Data types exported by `SYSTEM`
 Memory management

- Allocating and deallocating memory
- Memory access
 - Peeking and poking and addresses
- Logical operations
 - Bit twiddling
- Inline code
 - Why bother with a compiler?
- Miscellaneous
 - And all the rest...
- Reference
 - Module SYSTEM Reference

1.14 Data types

All data types imported from the pseudo-module SYSTEM must be qualified with the name of the module or an alias. For example, WORDSET must be referred to as SYSTEM.WORDSET.

The SET type in Oberon-A is a 32 bit entity. However, many Amiga data structures contain the equivalent of sets that are 8 and 16 bit entities. These smaller sets are represented by the BYTESET (8 bit) and WORDSET (16 bit) types exported by module SYSTEM. All the normal set operations may be performed on these types. The different set types are NOT compatible; sets of different types may not be mixed in expressions or assigned. Set constants have their types automatically adjusted by the compiler to conform to the type of set they being used with.

The operation of the LONG and SHORT standard procedures has been extended to deal with set type conversions. The \$P compiler switch must be set to FALSE to get access to these extensions. The LONG procedure will convert a BYTESET to a WORDSET and a WORDSET to a SET. The SHORT procedure will convert a SET to a WORDSET and a WORDSET to a BYTESET. This is the only supported method of mixing set types in assignments and expressions.

Module SYSTEM exports three anonymous types, BYTE, WORD and LONGWORD. These types are compatible with any other type with the same number of bits. Any 8-bit type (SHORTINT, CHAR, and BOOLEAN) may be assigned to a variable or parameter of type BYTE. In addition, any type may be passed to parameter of the type ARRAY OF BYTE. Any 16-bit type (INTEGER and WORDSET) may be assigned to a variable or parameter of type WORD. Any 32-bit type (LONGINT, SET, real types, pointers and procedures) may be assigned to a variable or parameter of type LONGWORD.

Three pointer types are exported: PTR, CPTR and BPTR. These are used as anonymous pointer types and are analogous to the LONGWORD type. Any Oberon pointer may be assigned to a variable or parameter of type PTR. Any CPointer may be assigned to a variable or parameter of type CPTR. In addition, a CPTR may be assigned to any variable or parameter of a

CPointer type. Any BPointer may be assigned to a variable or parameter of type BPTR. No other operations except comparisons with and assignment of NIL are allowed for these types.

The VAL function procedure is used to cause the compiler to treat an object of one type as if it had another type. This version of the compiler does not insist that the two types have the same size. This can cause unexpected problems with a big-endian processor like the MC68000. For example, if you convert a 32 bit type to a 16 bit type, you may end up accessing the `_upper_` 16 bits of the original object when you really wanted the `_lower_` 16 bits.

1.15 Memory management

The SYSTEM.NEW procedure is used to allocate a block of memory with an arbitrary size. Such a block does NOT have a type tag associated with it, so do not use this procedure to allocate a record structure through an Oberon pointer.

The DISPOSE procedure is used to explicitly free the memory associated with any pointer variable. Great care must be taken with this procedure, since it introduces the possibility of errors such as hanging pointers that Oberon is attempting to eliminate. The only valid use for DISPOSE is to free memory allocated to a CPointer or BPointer, or allocated using SYSTEM.NEW. DISPOSE makes sure that it has been passed a valid pointer and causes a processor trap to occur if it has not. It can be quite slow to execute in some circumstances (especially when freeing a pointer allocated in the middle of a large number of other allocations).

1.16 Memory access

The ADR procedure is used to find the run-time address of any variable or string constant. The result may be assigned to any CPointer. The BIND procedure is used to assign the address of a record variable to a CPointer variable while ensuring that the types of the two objects are compatible.

The BIT procedure is used to test an individual bit at a given memory location. Procedure GET is used to read a value at a given memory location while PUT is used to write one.

1.17 Logical operations

LSH, ROT, LOR, AND and XOR perform logical operations on integers. LSH is similar to ASH but performs a logical shift instead of an arithmetical shift (the difference is in the treatment of the sign bit). ROT performs a bitwise rotation of the argument. LOR performs a logical OR, AND a logical AND and XOR a logical exclusive-OR. Note that these operations do not change the type of the operand, unlike ASH which

promotes its parameter to a LONGINT.

1.18 Inline machine code

PUTREG is used to place a value in a specific CPU register. GETREG is used to read the value in a register. INLINE is used to insert machine code directly in the code buffer. It will output either a word or a longword, depending on the size of the type of the argument. INLINE will accept any number of parameters.

1.19 Miscellaneous

Procedure ARGS fetches the raw argument data passed to the program by AmigaDOS or Workbench. This will either be a pointer to a string containing the command line received from AmigaDOS, or a pointer to the Workbench startup message. Procedure ARGLEN is used to determine which it is. It returns -1 if the program was started by Workbench and the ARGS result should be interpreted as a message pointer. Otherwise it returns the length of the command line string and the ARGS result is a pointer to that string. See the implementation of module Args for an example of how to use these procedures.

Procedure SETCLEANUP is used to specify a procedure that is to be executed after the program has formally ended, either normally or as the result of a HALT or a processor trap. It returns the address of any procedure that had previously been specified elsewhere in the program. The new cleanup procedure should include a call to the previous procedure. The cleanup procedure must be declared as:

```
PROCEDURE <name> (rc : LONGINT); (* any parameter name will do *)
```

When a cleanup procedure is called, its parameter will contain the program's return code. If the program was terminated by a HALT statement, the parameter will be the value passed to HALT. Run-time errors may result in specific return codes: see module Errors for details. A procedure is not obliged to pass on the returnCode it receives if it calls another cleanup procedure.

Procedure MOVE is used to copy an arbitrary sequence of bytes from one memory location to another. It is able to deal correctly with overlapping blocks.

1.20 Module SYSTEM Reference

Function Procedures

v stands for a variable, x, y, a and n for expressions and T for a type. r stands for a register (0 ≤ r < 16).

Name	Argument type	Result type	Function
------	---------------	-------------	----------

ADR(v)	any	CPTR	address of variable v, or string constant v
AND(x, y)	x, y: integer type	larger type	logical AND
BIND(T, v)	T: CPointer type v: record type	T	address of variable v, asserting its type is a base type of T
BIT(a, n)	a: LONGINT n: integer type	BOOLEAN	Mem [a][n]
LSH(x, n)	x, n: integer type	type of x	logical shift
OR(x, y)	x, y: integer type	larger type	logical OR
ROT(x, n)	x, n: integer type	type of x	rotation
SIZE(T)	any type	integer type	size of T in bytes
VAL(T, x)	T, x: any type	T	x interpreted as type T
XOR(x, y)	x, y: integer type	larger type	logical exclusive OR

Proper Procedures

v stands for a variable, x, y, a and n for expressions and T for a type.

Name	Argument types	Function
ARGLEN(v)	v : LONGINT	v := length of command line, or -1 if the program started from Workbench
ARGS(v)	v : LONGINT	v := command line or workbench startup message
DISPOSE(v)	any pointer type	free memory allocated to v
GC	none	garbage collect memory
GET(a, v)	a: LONGINT v: any basic type	v := Mem [a]
GETREG(r, v)	r: register number v: any basic type	v := R[r]
INLINE(x1, ..., xn)	integer constant	insert x1 .. xn into code
MOVE(v0, v1, n)	v0, v1: any type n: integer type	assign first n bytes of v0 to v1
NEW(v, n)	v: any pointer type n: integer type	allocate block of n bytes and assign its address to v

PUT(a, x)	a: LONGINT x: any basic type	Mem [a] := x
SETCLEANUP(p,v)	p: parameterless procedure v: parameterless procedure variable	make p the global cleanup procedure; v := the previous cleanup procedure
SETREG(r, x)	r: register number x: any basic type	R[r] := x

1.21 Controlling the compiler with switches

The behaviour of the compiler is to some extent under programmer control. This control is exercised through compiler switches. ←

Compiler switches are embedded in comments and consist of a "\$" character, a single capital letter and a "+", "-" or "=" character. There must be no spaces between any of these characters. Any number of switches may appear in the same comment. The switches are boolean variables and their value is set by the trailing character, TRUE for "+", FALSE for "-" and the default value for "=".

Module switches must be specified before any imports or declarations and apply to the entire module. Global switches may be specified at any point in the source file and apply until changed or until the end of the module. Procedure switches apply to the first procedure body following the switch; at the end of the procedure body the switch returns to its default value. Note that it is procedure **body**, not procedure **declaration**; watch out for this with nested procedures.

The currently supported switches are:

Module switches

```
$P
  Portable code

$Z
  Zero variables
```

Global switches

```
$I
  Index checking

$L
  Addressing global variables

$N
```

Checking memory allocations

\$R
Range checking

\$T
Type checking

\$V
Overflow checking

Procedure switches

\$D
Copying open arrays

1.22 Copying open arrays

(* \$D+ enable copying of open arrays *)
(* \$D- disable copying of open arrays *)

\$D+ is the default. When TRUE, the compiler will generate code at the entry to each procedure to make copies of any open array value parameters. When FALSE, this code is suppressed. Use this switch for efficiency if you know an open array value parameter will be strictly read-only. Great care must be taken with this switch, since it effectively converts value parameters into variable parameters, bypassing the compiler's normal safety checks. The programmer MUST make sure not to make any assignments to such parameters, otherwise undesirable side effects such as writing over string constants may occur. This is a procedure switch.

1.23 Index checking

(* \$I+ enable array index checking *)
(* \$I- disable array index checking *)

\$I+ is the default. When FALSE, the compiler will suppress the generation of code to check that variables used to index arrays contain legal values. This is a global switch.

1.24 Addressing global variables

(* \$L- access global variables through A4 *)
(* \$L+ use absolute long addressing for global variables *)

\$L- is the default. When FALSE, the compiler generates code at the entry to each exportable procedure to set up the A4 register to point to

the module's global variables. The old value in A4 is saved and restored on exit from the procedure. This allows the most efficient access to global variables at the expense of several bytes (currently 10) of overhead per procedure. When TRUE, this code is suppressed and global variables are accessed with less efficient absolute 32-bit addresses. Set this switch to TRUE to minimise code size when you know that no global variables will be accessed in a procedure, OR IN ANY LOCAL PROCEDURE CALLED DIRECTLY OR INDIRECTLY BY IT. This is a global switch.

1.25 Checking memory allocations

(* \$N+ check for NIL result after call to NEW. *)
(* \$N- ignore NIL result from call to NEW. *)

\$N+ is the default. When TRUE, the compiler generates code after each call to the NEW or SYSTEM.NEW procedures to check that the memory allocation has succeeded. If not (the pointer returned is NIL), a processor trap is generated. When FALSE, the check is omitted and it becomes the programmer's responsibility to check that the allocation succeeded and take appropriate action if it did not. This is a global switch.

1.26 Portable code

(* \$P+ allow only portable code *)
(* \$P- allow non-portable code *)

\$P+ is the default. When TRUE, the compiler will only translate the Oberon-2 language as defined in the Oberon-2 Report. When FALSE, the compiler will recognise a number of language extensions (see

Language Extensions
) . This is a module switch.

1.27 Range checking

(* \$R+ enable range checking *)
(* \$R- disable range checking *)

\$R+ is the default. When TRUE, the compiler will generate code to check that variables contain legal values for particular operations. This includes the IN, INCL and EXCL set operations and type conversions using CHR and SHORT. When FALSE, this code is suppressed. This is a global switch.

[Not all of these checks have been implemented as yet. FJC]

1.28 Type checking

(* \$T+ enable type checking *)
(* \$T- disable type checking *)

\$T+ is the default. When FALSE, the compiler will suppress the generation of code to perform run-time type checks on variable record parameters and pointers. This is a global switch.

1.29 Overflow checking

(* \$V+ enable overflow checking *)
(* \$V- disable overflow checking *)

\$V+ is the default. When FALSE, the compiler will suppress the generation of code to check for arithmetic overflow at run-time. This is a global switch.

1.30 Zeroing variables

(* \$Z- suppress zeroing of variables *)
(* \$Z+ zero all global and local variables *)

\$Z- is the default. When TRUE, the compiler generates code in the main body of the module and on entry to each procedure to fill all variables with zeroes. This guarantees that variables have a predictable initial value. Numeric variables are set to zero; booleans are set to FALSE; pointers are set to NIL; sets are empty, ie - {}. When FALSE, this code is suppressed and variables have their values undefined. Why is the default set to FALSE? Because in the language report the initial values of variables are undefined and if you write programs that make assumptions about initial values they will not be portable. The default forces you to explicitly state that your program is relying on non-standard behaviour from the compiler. This is a module switch.

1.31 Using the garbage collector

Oberon-2 was designed under the assumption that programs written in it would be running in an environment that provided automatic garbage collection of memory. This is the reason why it has a NEW standard procedure but no DISPOSE. The Amiga's operating system does not provide this facility, so Oberon-A implements a garbage collector in the run-time support code linked with every program. This garbage collector must be used carefully, as it has the potential to free memory that is still in use.

The garbage collector is invoked by calling the GC procedure in the pseudo-module SYSTEM. When called, it works in two phases: a mark phase and a scan phase. During the mark phase it traces all the global

pointer variables and marks the memory they point to. If the marked memory contains other pointers, either as record fields or array elements, these are also traced and marked. When the mark phase is completed, the scan phase processes a list of all the memory blocks allocated by the program, unmarking any marked blocks and freeing all unmarked blocks.

The point in the program at which the garbage collector is called is very important. The mark phase can only trace memory accessible from GLOBAL pointer variables. LOCAL pointer variables inside procedures cannot be traced. If such local variables are still active, the memory allocated to them will be freed, almost certainly leading to a crash. To avoid this, the programmer must ensure that the garbage collector is only called at a point in the program where it is guaranteed that there are no active local pointer variables. An ideal place for this would be in the program's main event loop (if it is a GUI program). A counter variable should be used to limit the frequency at which the collector is activated; activating it every cycle of the loop would bring the system to a halt.

Another danger comes from using the SYSTEM.DISPOSE procedure. If there is more than one reference to memory freed with this procedure, the garbage collector will be tricked into believing that the memory is still allocated, causing it to write all over memory it doesn't own. If you cannot guarantee that you know of all references to a dynamically allocated variable and have assigned NIL to all of them, DO NOT USE SYSTEM.DISPOSE. Assign NIL to any global pointer variable you are finished with, and trust the garbage collector to handle any other references. This kind of bug is very difficult to track down. When it happened to the compiler, it took almost a week to find (and 30 seconds to fix). Debuggers were useless, as they were being crashed by random memory writes. You have been warned.

A number of library modules distributed with Oberon-A allocate memory in their operations. For the reasons given above, most do not call SYSTEM.DISPOSE. Module Files is a notable example, allocating from one to four 1K buffers for every file opened. If you use such modules intensively, you are more or less obliged to call the garbage collector periodically to avoid running out of memory.

Garbage collection applies only to Oberon pointers. CPointer and BPointer variables are not traced and the garbage collector ignores them. If you use NEW or SYSTEM.NEW to allocate memory to such pointers, you should use SYSTEM.DISPOSE to free them. This is equivalent to using C's malloc() and free() functions.

You are not forced to use either the garbage collector or SYSTEM.DISPOSE. Any memory allocated by a program that is not explicitly (with SYSTEM.DISPOSE) or implicitly (with the garbage collector) freed, will be automatically returned to the system when the program ends. This happens even if the program crashes due to a processor trap or is summarily terminated with HALT or ASSERT. It also applies to memory allocated to CPointer and BPointer variables with NEW and SYSTEM.NEW. IT DOES NOT APPLY TO MEMORY ALLOCATED WITH THE AMIGA MEMORY ALLOCATION FUNCTIONS. The run-time system cannot track such memory and if it is not explicitly freed it will remain allocated and cause a memory leak.

1.32 Handling run-time errors

The compiler generates code fragments to check for a number of errors that may occur at run-time. These include arithmetic overflows, failed type guards, array index errors, etc. They can be enabled and disabled with compiler switches; they are all enabled by default. Typically run-time errors produce a processor trap with a TRAP or TRAPV instruction. The run-time support code built into every Oberon-A program contains a trap handler which intercepts all compiler-generated traps and several others such as divide-by-zero. The default trap handler has the same effect as a HALT statement, causing the program to terminate. Any cleanup procedures installed with SYSTEM.SETCLEANUP will be executed and all memory allocated with NEW or SYSTEM.NEW will be freed. The return code will be set to the trap number + 100.

Module Errors gives an example of a cleanup procedure which checks the return code and puts up a requester describing the error. This example should give you enough information to write your own replacement, or a supplementary procedure that catches return codes it doesn't understand. If you know what you are doing, you could install your own trap handler through the tcTrapCode field in the program's Task structure. See the Amiga RKM for details.

1.33 Error reports produced by the compiler

Any errors detected by the compiler are listed in the file "Oberon.errors". This will be placed in the T: directory if possible, otherwise in the current directory. Errors are listed one per line. Each report takes the format:

```
line <line#>, col <column#>: err = <error#>
```

Lines and columns are numbered starting at 1. The meaning of each error number is listed in the file ErrorCodes.doc.

1.34 Implementation of basic types

The Oberon Report leaves the precise format and size of most basic types up to individual implementations. The relevant data for Oberon-A are:

Type	Size	MIN	MAX
----	----	---	---
SHORTINT	8 bits /1 byte	-128	127
INTEGER	16 bits/2 bytes	-32768	32767
LONGINT	32 bits/4 bytes	-2147483648	2147483647
REAL	32 bits/4 bytes	-9.22337177E18	9.22337177E18
LONGREAL	32 bits/4 bytes	-9.22337177E18	9.22337177E18
CHAR	8 bits /1 byte	0X	255X
BYTE	8 bits /1 byte	0	255
SYSTEM.BYTESET	8 bits /1 byte	0	7
SYSTEM.WORDSET	16 bits/2 bytes	0	15

SET	32 bits/4 bytes	0	31
Pointers	32 bits/4 bytes	N/A	N/A

Note that REAL and LONGREAL are identical in this implementation. They both conform to the Motorola Fast Floating Point standard. In a future version, LONGREAL will be re-implemented as an IEEE double-precision real. REAL may also be re-implemented as an IEEE single-precision real. I will be guided in this by the compiler's users.

1.35 Limits built in to the compiler

- * The compiler cannot evaluate constant expressions that contain REAL or LONGREAL values. You can still have real literals, just no arithmetic.
- * The code generated for a module cannot be greater than 32K in size. This limit is inherent in the addressing model used for generating machine code. Split large modules into smaller pieces.
- * No more than 4000 bytes of string literals or 1000 bytes of type tag data may be generated for a module. These limits can be increased by changing two constants in module OCC and recompiling it; the absolute maximums are 32K.
- * There is a limit of about 80K on the size of the buffer used to store names. This is dynamically allocated and grows as needed, so there is no wasted RAM. The most I have seen used for this is about 60K.
- * No more than 32K of local variables can be declared for a procedure. What do you mean you want more? Use dynamic allocation.
- * There is no limit on the size of a module's global variables but variables more than 32K from the module's variable base will be less efficient to access.
- * An array type may not have more than 32K elements (however, it may be larger than 32K bytes). This limit will disappear in a future version.
- * Identifiers and string literals cannot be more than 255 characters long. This is primarily a limit imposed by module OCS, but increasing it may affect assumptions made in other parts of the compiler. The original limit for identifiers was 31 characters and this may still lurk in dark corners of the source code. Of course, if you need identifiers longer than this, stop writing variable names in German :-). Module names are limited to 26 characters. This limit is imposed by AmigaDOS.
- * String literals longer than 1 character cannot be aliased if they are imported from another module. By this I mean, you cannot declare a constant such as:

```
CONST Alias = AnotherModule.StringConstant;
```

where StringConstant is a string literal longer than 1 character. This

limit will probably disappear in a future version. It will happen quicker if people complain :-).

- * The register allocation strategy implemented by the compiler prevents the use of Amiga shared library functions that pass parameters in the A4 register. Those that use the A3 register may also pose some difficulties. If the actual parameter is an expression rather than a simple variable, you may need to assign it to a temporary variable and pass the temporary to the parameter. This will be fixed in a future release.
- * There are a number of arbitrary limits placed on the number of objects such as exported types, imported modules and the like. These limits allow the use of arrays for internal data structures, which are much more efficient than dynamically allocated lists. Most of these limits have been greatly increased from those in the Ceres compiler. If you still manage to exceed such a limit, a compiler error will be reported and you should easily be able to determine which constant to increase to get around it.
- * The compiler needs at least 10000 bytes of stack and up to 500K of free RAM to run.

1.36 Who is responsible for THIS?

OC was ported to the Amiga by Frank Copeland. It is based on a compiler written by Niklaus Wirth.

All bug reports, suggestions and comments can be directed to:

Email : oberon@wosname.apana.org.au

Snail Mail :

Frank J Copeland
PO BOX 236
RESERVOIR VIC 3073
AUSTRALIA

Remember the J. It saves a lot of confusion at my end :-).

1.37 Reporting bugs and suggestions

This version of OC is a beta-test version. That means that it is basically complete, but has not been rigorously tested. Bug fixes and suggestions from users of this version will be incorporated in future versions. You are encouraged to report any and all bugs you find, as well as any comments or suggestions for improvements you may have.

Before reporting a suspected bug, check the file ToDo.doc to see if it has already been noted. If it is a new insect, clearly describe its behaviour including the actions necessary to make it repeatable.

Indicate in your report which version of OC you are using. Include an example of a program or short fragment of code that demonstrates the bug.

I am especially interested in the following areas:

- * Compatibility with different versions of the Amiga hardware and operating system. So far OC has only operated on a stock A500 with AmigaDOS 2.05 and a 20MB hard disk.
- * How good/useful/helpful/complete the documentation is.
- * How suitable OC is for use by programmers with varying levels of experience, from beginners to hackers.
- * Departures from the language specification.
- * Extensions to the language supported by the compiler.
- * Memory management. I am unable to use Enforcer or similar utilities on my A500, so I would like people who can to report any Enforcer hits they get. I am also concerned about possibly excessive memory fragmentation caused by the run-time memory allocator.

1.38 Who did what and why

OC is a port of a compiler written for the Ceres workstation by Niklaus Wirth. The book "Project Oberon" written by Wirth and Jürg Gutknecht contains a description of this compiler and the full source code for it. The original source can also be obtained by anonymous ftp from [neptune.inf.ethz.ch](ftp://neptune.inf.ethz.ch). Many thanks to Professor Wirth for making this source code available.

The machine code generator for early versions of the compiler was a part of part of Charlie Gibb's A68K assembler. This code is no longer part of the compiler, but it was extremely useful in the early stages of development and debugging.

Part of the run-time library (the 32 bit arithmetic) is taken from the Sozobon C compiler and is:

Copyright (c) 1988 by Sozobon, Limited. Author: Johann Ruegg

All the source code for OC was created using Matt Dillon's DME editor.

1.39 Release history

0.0 The initial port to the Amiga, written in Modula 2 and compiled by the Benchmark compiler. Implemented the Oberon dialect. Never released. Started in February 1993.

-
- 0.1 The initial conversion from Modula 2 to Oberon, compiled by the v0.0 compiler. Never released.
 - 0.2 - 0.3 Bug fixes and upgrades. Never released.
 - 1.0 Start of revision control. Upgrades and bug fixes. Never released.
 - 2.0 First public release. Compiler upgraded to Oberon-2. Released in May 1994.
 - 3.0
 - * Changed command line arguments:
 - Options now must come first;
 - Multiple filename arguments allowed.
 - * Batch compiles implemented.
 - * OLIB: is now the default symbol file search path.
 - * Error files are output in the current directory with the name "<module>.err".
 - * [bug] Enforcer hit caused when no DST parameter was specified
 - * [bug] Same error code (#228) used for different errors.
-