

# Release Notes — Gardens Point Component Pascal

## Version 1.2.0 for .NET (*September 2002, RTM version*)

### 1. Introduction

Gardens Point Component Pascal (*gpcp*) is an implementation of the Component Pascal Language, as defined in the Component Pascal Report from Oberon Microsystems. It is intended that this be a faithful implementation of the report, except for those changes that are explicitly detailed here. Any other differences in detail should be reported as potential bugs.

The distribution consists of four programs, and a number of libraries. The programs are the compiler *gpcp*, the make utility *CPMake*, a module interface browser tool *Browse*, and a tool for producing interfaces from assemblies written in other .NET languages *N2cps*. There will be other utilities added later.

The compiler produces either Microsoft.NET intermediate language (“MSIL”) or Java byte-codes as output. The compiler can be bootstrapped on either platform. These release notes refer to the Microsoft.NET platform.

There are a number of syntactic extensions to the Component Pascal language accepted by the compiler which are introduced to allow interworking with the native libraries of the underlying platform. The guiding philosophy in such cases is to not significantly extend the semantics of the constructs that form part of Component Pascal, but rather to provide syntax for accessing features of other languages, which have no direct counterpart in Component Pascal.

### 2. Overall Structure

#### 2.1 Input and Output files

In normal usage the compiler creates either three or four output files for every source file. If the file **Hello.cp** contains the module “Hello”, and is compiled, then the output files will be **Hello.cps**, **Hello.il**, and either **Hello.dll** or **Hello.exe**. The “\*.cps” file is the symbol file which contains the meta-information that describes the facilities exported from the module. The “\*.il” file contains the MSIL intermediate language representation of the program. The program executable will be “\*.exe” if the program contains an entry point (i.e. if the module imports **CPmain**), otherwise the compilation will create a dynamic link library “\*.dll”. All of these files are created in the current directory. If a listing file is created it will have extension “\*.lst”.

Be aware that the stem name of the output files comes from the *module* name, not from the source-file name. Thus if module “Foo” is in source file “Hello.cp” then all of the output files will have stem name “Foo”.

It is possible to invoke the compiler so as to produce just the intermediate language file, and then invoke the intermediate language assembler manually. The assembler **ilasm** may then be used to produce any of its possible output formats.

## 2.2 Invoking the compiler

The compiler is invoked from the command line using the command

```
$> gpcp [options] files↵
```

where options include

<b>/copyright</b>	display the copyright notice
<b>/help</b>	emit this usage prompt
<b>/list</b>	create an output listing if there are errors (default)
<b>/list+</b>	always create an output listing
<b>/list-</b>	never create an output listing
<b>/debug</b>	emit debugging symbols (default)
<b>/dostats</b>	emit timing and other statistics
<b>/hsize=NNN</b>	set hashtable size $\geq$ NNN (0 .. 65000)
<b>/ilasm</b>	use <b>ilasm</b> even with <b>/nodebug</b>
<b>/noasm</b>	produce a symbol file, but no il
<b>/nocode</b>	create il output, but do not assemble
<b>/nodebug</b>	do not emit debugging symbols
<b>/nosym</b>	produce no output files, not even a symbol file
<b>/nocheck</b>	produce code without arithmetic overflow checks
<b>/verbose</b>	chatter on about progress during compilation
<b>/version</b>	emit version information
<b>/warn-</b>	suppress warning messages from the console
<b>/nowarn</b>	same as <b>/warn-</b>
<b>/target=xxx</b>	emit assembler output for platform “xxx”
<b>/special</b>	used for creating symbol files for foreign interfaces
<b>/strict</b>	disallow the use of non-standard language constructs
<b>/symdir=xxx</b>	place symbol files in directory “xxx”
<b>/bindir=xxx</b>	place PE (and pdb) files in directory “xxx”

Any number of files may be added in a white-space separated list. The compiler also accepts Unix-style comments starting with ‘-’. In the *JVM* version the ‘-’ form is the expected default.

## 2.3 Target choice

The compiler may choose its output language at runtime. The default output when running on the *.NET* platform is *.NET* assembler (.il) for the *.NET* virtual object system. The recognized options are –

<b>/target=net</b>	this is the default <i>.NET</i> virtual object system format
<b>/target=jvm</b>	this causes Java byte codes to be emitted
<b>/target=dcf</b>	this chooses the Gardens Point “d-code” form

The Java output option produces either jvm class files directly, or produces assembly language files for the “Jasmin” byte code assembler.

The dcf format is not yet available, but is intended to access the Gardens Point native code generators on all the platforms for which Gardens Point Modula-2 is implemented.

### Output files

Running the compiler with the **/nosym** flag causes the input files to be parsed and type-checked, but no output files are created except possibly a listing file.

If the compiler is run with the **/noasm** flag, the input files are parsed and type-checked, and a symbol file is produced for each input file. No assembly language or program executable file output is produced however.

If the compiler is run with the **/nocode** flag, the input files are parsed and type-checked, and a symbol file and one MSIL assembly language file is produced for each input file. No executable files are produced in this case.

If the compiler is run without any flags, the input files are parsed and type-checked, and a symbol file, and a program executable (PE) file (either .DLL or .EXE) is produced for each input file. If the default **/debug** flag is in effect then a textual IL file (extension .il) is produced and the PE file is created by the **ilasm** tool. If the **/nodebug** flag is in effect then a PE-file is directly produced using the **PEAPI** component. This behavior may be overridden by use of the **/ilasm** flag, which causes textual IL to be produced, even in the presence of the **/nodebug** flag.

### Output files with **/target=jvm**

If the compiler is run with the **/target=jvm** flag, the input files are parsed and type-checked, and a symbol file and one or more class files will be produced. These class files are written directly, and do not require the installation of Jasmin.

If, in addition, the **/nocode** flag is used, then Jasmin assembly language (\*.j) files will be produced, but the assembler will not be invoked.

If, instead of **/nocode** the **/jasmin** flag is added, Jasmin assembly language files are produced for each input file. Following this, the Jasmin assembler will be automatically invoked to create the corresponding class files. Because a separate process needs to be created for each invocation of Jasmin, this is quite slow.

## 2.4 Overflow checking

Ordinarily the compiler produces code that performs arithmetic overflow checks on all operations. Narrowing assignments (such as assigning a long value to an integer variable) are also range checked. Compiling with the **/nocheck** option removes these checks. There is a very small speed gain if checks are turned off. Checks may also be turned off on a per-procedure basis, as described below.

## 2.5 Listing output

The compiler, by default, produces a listing file only if there are compile-time errors or warnings. It is possible to force the compiler to produce a listing, using the **/list+** option. Equally, it is possible to prevent the creation of a listing file even if there are errors, by using the **/list-** option.

The listing file contains the complete listing of the program, with four digit line numbers prepended. Errors are reported in the following format –

```

1  MODULE BarMod;
2      IMPORT
3          FooMod;
4      TYPE
5          Bar* = POINTER TO ABSTRACT RECORD (FooMod.Foo) END;
****      ^ Only ABSTRACT basetypes can have abstract extensions
6          i,j,k : INTEGER
7          END;
8  END BarMod.

```

## 2.6 Statistics output

If the compiler is invoked with option `/dostats` then compile time statistics are produced. Here is an example for the program *Browse*.

```

C:\gpcp\work> gpcp /dostats Browse.cp
#gpcp: created Browse.exe
#gpcp: <Browse> No errors
#gpcp: net version 1.1.4 of 25 December 2001
#gpcp: 2165 source lines
#gpcp: import recursion depth 3
#gpcp: 796 entries in hashtable of size 4099
#gpcp: import time      70mSec
#gpcp: source time     110mSec
#gpcp: parse time      231mSec
#gpcp: analysis time   60mSec
#gpcp: symWrite time    20mSec
#gpcp: asmWrite time   210mSec
#gpcp: assemble time   391mSec
#gpcp: total time      1092mSec

```

The meaning of the values written to the console is as follows.

- The compiler imports symbol files in dependency order, if necessary. The maximum recursion depth for this example turned out to be 3.
- The size of the hash-table, and the number of entries used is shown
- Import time is the time to read and process metainformation for all imports. In the example *Browse* imports much of the compiler data structures.
- Source time is the time to read the source file into the internal buffer.
- Parse time is the time to parse the buffer, create the syntax tree and resolve all identifiers.
- Analysis time is the time to do type checking, and dataflow analysis.
- SymWrite time is the time to write out metatdata to the symbol file.
- AsmWrite time is the time to write out the assembly language (il) output.
- Assemble time is the time taken to spawn a new process and run `ilasm` (or `jasmin` if `/jasmin` has been set). Assemble time is always zero if the direct to PE output path is selected by `/nodebug`.

## 2.7 Setting the hash table size

The compiler uses closed hashing internally, with a default number of identifiers of 4099 in the current version. It is possible to increase the number of entries by means of the `/hsize=NUMBER` option. Numbers up to 66000 are meaningful to the program.

If the hash table overflows, the compiler gives an error message, with a hint to increase the size. There is a example program with the distribution that creates a program that

will break the compiler, so that users may test this feature. The compilation fails with the default table, but succeeds with `/hsize=5000`.

## 2.8 Choosing the Output Directories

By default all output files are created in the current directory. This behavior may be overridden with the new options `/bindir` and `/symdir`. The symbol file is placed in the directory specified by the option –

`/symdir=target-directory`

Note carefully that if a target directory is chosen that is not on the “CPSYM” path then *gpcp* will not be able to find the symbol files automatically.

Program executable directories, and debug files in the case that debugging symbols are being created may be placed in a specified directory using the –

`/bindir=target-directory`

option.

If the JVM target has been chosen then the “symdir” option still applies, but “bindir” option does not. Instead, the root of the output class file hierarchy may be specified using a syntactically similar “clsdir” option.

## 2.9 The Make utility

The compilation process with Component Pascal guarantees type safety across separately compiled module boundaries. Since interface meta-information resides in the symbol files which *gpcp* creates, modules must be compiled in an order that respects the partial order induced by the global importation graph. For complex programs, this may be difficult to determine manually.

The utility *CPMake* reads symbol files, and if necessary source files, in order to determine a valid order of compilation. The syntax for invocation is –

`$> CPMake [options] moduleName←`

The module name may be given with or without a file-extension, but must be the name of a module which imports *CPMain*, that is, it must be a base module. The module name given to *CPMake* is case sensitive.

When source files of a program have been modified in general only a subset of the modules have to be recompiled. *CPMake* is able to work out which modules must be recompiled by checking the date stamps on the files, and also checking the module hash-keys (“magic numbers”) in the symbol files. If a module has been edited, but the public interface of the module has not changed a recompilation should compute a new magic number that is the same as that expected by any previously compiled, dependent modules. In this case *CPMake* detects that the dependent modules are still consistent and do not require recompilation. This “domino-stopping” feature of the program ensures that a conservative minimum of modules are recompiled.

The options accepted by the program are exactly the options accepted by *gpcp*, except that an additional option `/all` forces compilation of all modules in the local directory irrespective of date stamps and magic numbers. (If you use *CPMake* to bootstrap the compiler, be aware that *ilasm* will fail if the output would overwrite the file of a loaded assembly. This means that you cannot bootstrap the compiler using the compiler in the same directory, unless you use `/nocode` and then invoke *ilasm* manually.)

## 2.10 Module Interface Browser

The program **Browse** reads the symbol file of a module and displays the public interface. This public interface is shown in a form similar to a Component Pascal module. This “module” shows all the types, variables and procedures that are exported from the specified module. Only the exported fields of record types are shown. Any exported procedures are shown as procedure headers only. The output from **Browse** is **not** a proper Component Pascal module and will not compile using **gpcp**. It simply shows all of the identifiers that may be imported and used by a client module.

This program is invoked with the command

```
$> Browse [options] moduleName↵
```

The symbol file extension **.cps** may optionally be included in *moduleName*.

As with **gpcp**, any number of files may be added in a white-space separated list. The **Browse** program sends its output to the console by default, and has the following options:

<b>/all</b>	browse this and all imported modules
<b>/full</b>	display full foreign names
<b>/file</b>	write output to the file <i>&lt;moduleName&gt;.bro</i>
<b>/html</b>	write html output to the file <i>&lt;moduleName&gt;.html</i>

The **-all** option produces output for all of the modules on the global imports graph of the specified module. The **/full** option is only meaningful for FOREIGN modules where the output from **Browse** will include the full external names for all procedures. The default for **Browse** is to only display the internal (Component Pascal) names. See Section 7 for more on Foreign Language Interfaces. The **/file** option sends the output to the file *<moduleName>.bro* instead of to the console. The **/html** option produces hyperlinked html text in the file *<moduleName>.html*.

## 2.11 Symbol File Generator N2CPS

This program generates symbols files corresponding to *.NET* assemblies. Taken together with the **Browse** tool, this makes the libraries of the *.NET* framework accessible to Component Pascal users. Usage is -

```
$> N2cps [options] {assemblyName}↵
```

where current options are **-v** for “verbose” and **-w** for “warnings”. Each specified assembly will produce one or more symbol **\*.cps** files. For example, the main system library assembly **mscorlib** will produce **mscorlib\_System**, **mscorlib\_System\_Reflection** and so on. Until the final release of *.NET* it may be prudent to ensure that you have the *NetSystem* symbol files that exactly correspond to the version of *.NET* that you have installed. Go to the `gpcp\source\N2CPS` directory and run first “*NetClean*” then “*NetMake*”.

## 3. Lexical Issues

### 3.1 Non-standard Keywords

In order to provide facilities for the foreign language interface there are a total of six new keywords defined. These are all upper case names and cannot be used as program identifiers.

<b>DIV0</b>	- an additional arithmetic operator (C integer division)
<b>ENUM</b>	- used in dummy foreign modules in the <i>.NET</i> system
<b>INTERFACE</b>	- used in dummy foreign modules for defining interfaces
<b>REM0</b>	- an additional arithmetic operator (C integer remainder)
<b>RESCUE</b>	- used to mark a procedure-level exception catch block
<b>STATIC</b>	- used to declare static features in dummy foreign modules.
<b>EVENT</b>	- used to declare multicast delegate type for <i>.NET</i> events.

Only **DIV0**, **REM0**, **EVENT** and **RESCUE** may be used in normal programs, the remainder are used in dummy foreign definition modules.

The following new predefined identifiers have been added. These can be redefined, but not at the outer lexical level. Definitions for these procedures are given below.

<b>MKSTR</b>	- function to convert a CP “string” to the native string type
<b>BOX</b>	- make a dynamically allocated copy of record or array
<b>THROW</b>	- procedure that (re)throws a native exception object
<b>TYPEOF</b>	- fetch the runtime type descriptor, for reflection etc.
<b>REGISTER</b>	- attaches a procedure to a ( <i>.NET</i> ) multicast delegate
<b>DEREGISTER</b>	- detaches a procedure from a multicast delegate

There are some other predefined identifiers used in the extended syntax, but these are “context sensitive markers” and do not prevent the same names being used for program identifiers.

✓ *Remember, if you use any of these non-standard keywords or procedures, your program source will not be portable to other implementations of Component Pascal.*

### 3.2 Common Language Specification names

Fully qualified names in the Common Language Specification of *.NET* (CLS) comprise four parts.

<b>Assembly</b> name	- this specifies the dll in which the class will be found
<b>Namespace</b> name	- this specifies the namespace of the class
<b>Class</b> name	- the class name
<b>Feature</b> name	- the field or method name.

An example might be

**[mscorlib]System.Exception::ToString**

where **mscorlib** is the assembly name, **System** is the namespace, **Exception** is the class name, and **ToString** is a method name.

In this version of *gpcp*, the compiler produces one assembly per module, and one namespace per module. Both the assembly and the namespace names are the same as the module name. Thus a type-bound procedure called **isString()** bound to the type **UnaryX** in module **ExprDesc** would have the CLS name –

**[ExprDesc]ExprDesc.UnaryX::isString**

Procedures and variables at the module level are declared in the *CLS* as belonging to a synthetic “class” that contains only static data and code. This “**implicit static class**” has the same name as the module. Thus variable “x” in module Foo will have the somewhat boring *CLS* name

**[Foo]Foo.Foo::x**

Users of the compiler should almost never have to deal with explicit *CLS* names.

If you do browse the assembler output of the compiler, you will notice that almost all names are escaped with single quotes like ‘this’. This is done to avoid clashes with the many names that are reserved in the assembler.

All aspects of the default naming scheme may be overridden, if required. Such a necessity might arise if the Component Pascal must interface with a framework that has particular naming patterns hardwired in. The details of the mechanisms for overriding are given in Appendix B.

### 3.3 Identifier syntax

The identifier syntax for Component Pascal allows arbitrary use of the underscore (low-line) character. There is a further extension that is specific to the foreign language interface of *gpcp*.

Occasionally, names that are imported from foreign modules will happen to clash with CP reserved words. In this case, we may escape the reserved word detection by starting the identifier with the “back-quote” character ```. Thus, if an imported module has (say) a class with a field named “**IF**”, then the field may be referenced as ``IF` in the source of your program. You may not *define* identifiers using this escape mechanism, except in foreign definition modules. You may however *refer* to imported identifiers using this mechanism.

It may be important to know that the back-quote is stripped at the time that the program is scanned. The presence of the escape simply suppresses the usual check for reserved identifiers that normally follows identifier scanning. Thus the back-quote is not used during any name matching of identifiers. A curious result of this strategy is that if a program escapes an identifier that does not need it, the escaped and non-escaped identifiers will refer to the same name.

## 4. Semantic Issues

### 4.1 DLLs and EXEs

The compiler can produce either stand-alone executables (**.exe** files) or dynamic link libraries (**.dll** files). Executable files must have an entry point known to the runtime as **‘CPmain’**, optionally taking an array of strings as parameters.

If the source file contains the import of the special name **CPmain**, then an executable file is produced as output. In this case the module body becomes **.CPmain**, and begins with a hidden call which saves any command line arguments so that they may be accessed by calls to the *ProgArgs* library.

If the source file does not import **CPmain**, then the module body becomes the “class constructor” which is executed at the time that the dynamic link library is loaded on demand.



If the compiler is run with the `/nocode` option, then only the assembler (`.il`) file is created. In this case the assembler *ilasm* may be invoked so as to create either a `.dll` or an `.exe` file using the command `ilasm /DLL` or `ilasm /EXE`. Of course, it is an error to try to create an executable file if the source does not contain an entry point.

## 4.2 Unimplemented constructs

There are a small number of constructs that are unimplemented or restricted in this release of the compiler. These are –

- Procedure variables
- Module finalizers

Both of these features were implemented in a prototype version of the compiler. Procedure variables have been implemented, but with some restriction in the current release. Arbitrary procedures of matching type may be assigned to procedure variables, and called in the usual way. However *assignment* of procedure variables is only permitted if the two sides of the assignment have the same type. That is, assignment of procedure values other than literal procedures requires *name compatibility*, rather than the specified *structural compatibility*. This restriction will be removed in the next major release.

Non-local variable access is permitted in an unrestricted way since release 1.1.6. However, in the case of reference (VAR) parameters of unboxed type that are accessed from within nested procedures the semantics of parameter passing is modified because the actual parameters are passed by copying rather than by reference. The compiler gives an explicit warning in these unusual circumstances. Component Pascal types that are unboxed in the .NET implementation are scalar values and record types that are not extensible, do not extend another type, and are not defined as the anonymous bound type of a pointer type.

## 4.3 Additional Arithmetic Operators

The usual arithmetic operators **DIV** and **MOD** in Pascal-family languages have well defined semantics that are different to the division and remainder operators of implementations of C-family languages. In Component Pascal the operators **DIV** and **MOD** are defined as follows –

$$(i \text{ **DIV** } j) \times j + (i \text{ **MOD** } j) = i$$

$$i \text{ **DIV** } j = \lfloor i/j \rfloor; \text{ where } i, j \text{ are integers, and } i/j \text{ denotes real division.}$$

Notice that **DIV** always rounds toward negative infinity unlike most C-language implementations (which normally round towards zero). The Pascal operators are mathematically preferred, but in case the alternative semantics are required for compatibility reasons, *gpcp* introduces alternatives. **DIV0** denotes integer division with rounding toward zero, while **REM0** denotes the corresponding remainder operation.

☑ *Remember, if you use these non-standard operators, your program source will not be portable to other implementations of Component Pascal.*

#### 4.4 Semantics of the WITH statement

The semantics of the WITH statement have been slightly modified so as to strengthen the guarantees on the properties of the selected variable. In the code –

```
WITH x : TypeTi DO
    ... <guarded region>
| x : TypeTj DO
    ... <guarded region>
END;
```

the variable  $x$  is asserted to have the specified type throughout the so-called guarded region. The base language guarantees that the type of the selected variable cannot be widened in the guarded region, but might possibly be narrowed. In *gpcp* the selected variable is treated as a constant, and neither the type nor the value can be modified either directly or indirectly. Any attempt to do so attracts a compile-time error message.

#### 4.5 Implementing foreign interfaces

Component Pascal types may extend classes from the *.NET CLS*. Types which extend *CLS* classes may also declare that they implement interfaces from the *CLS*. The syntax extension to access this feature is –

```
RecordDecl ::-      RECORD [BaseType] [Fields] END;
BaseType ::-        “[“ QualifiedIdent { “+” QualifiedIdent } “]”
```

The first qualified identifier, as in the *Report*, is the class that is extended by the type being defined. Any additional qualified identifiers are the names of interfaces that the type promises to implement. The compiler checks that this contract is honored. In the case that interfaces are implemented, the base type may be left blank, or may be explicitly set to **ANYREC**.

The semantics of type casts are also relaxed whenever a reference is cast to an imported interface type. For non-interface types many erroneous casts can be detected at compile time, but for interfaces no cast of an object of a foreign type can be rejected at compile time.

It is not possible to *define* interface types in Component Pascal.

#### 4.6 EVENT types

Event types are declared in *gpcp* with the same syntax as procedure types, but with the keyword **PROCEDURE** replaced by **EVENT**. Events are implemented as multicast delegate type in the *.NET* framework. If variables are declared to be of some event type, then it is possible to use the new builtin procedures **REGISTER** and **DEREGISTER** to register or deregister callbacks on the multicast delegate.

The usage for registering a callback is

```
REGISTER( target-location , callback-method ) ;
```

The target location is the designator of the location of event type. The specification of the callback method has two forms. If a proper procedure is to be added, then the procedure name is used. If the callback is to invoke a particular method on a particular object, then the syntax *object.method* is used. This works for any typebound procedure

in Component Pascal. The usage for deregistering a call is syntactically the same, but uses the new builtin procedure **DEREGISTER**.  
A callback may be registered multiple times. The delegated calls are made in order of registration.

#### 4.7 Unsigned Byte Type

The 8-bit type used in the .NET *Common Language Specification* (CLS) is an unsigned type. If Component Pascal is to be a full consumer of CLS libraries then it must be possible to declare variables and fields of such types in Component Pascal programs. In order to facilitate this a new built-in type **UBYTE** has been introduced in version 1.2 of *gpcp*. Values of the type may be assigned to variables of larger integral types as required. However, if values of this type are assigned to locations of the signed 8-bit type **BYTE** a runtime check is required. Similarly if values of any signed type are assigned to a location of unsigned byte type an explicit narrowing is required, using the new built-in function **USHORT**( ).

#### 4.8 Runtime type descriptors

A new function in this release returns the runtime type descriptor. This allows easy access to the facilities of the reflection libraries. The function is overloaded, and has the following signatures –

```
PROCEDURE TYPEOF(typename) : RTS.NativeType;  
PROCEDURE TYPEOF(IN s : anytype) : RTS.NativeType;
```

If the target is .NET, then *NativeType* will be `System.Type` on the underlying runtime. If the target is the JVM, then the return value will be `java.lang.Class`

The first version takes any type name as actual parameter. The second version takes an actual parameter that is any variable designator. If the type of the designator is statically known (perhaps because it denotes an object of an inextensible type) then the compiler resolves the reference and no call is made to the runtime function `System.Object::GetType`( ).

#### 4.8 Additional built-in functions

There are four additional built-in functions added to the implementation. One allows convenient access to the underlying native string object type. The signature is –

```
PROCEDURE MKSTR(IN s : ARRAY OF CHAR) : RTS.NativeString;
```



*Note that it is never necessary to use MKSTR when passing a literal string to a formal parameter of native string type. In the literal case the compiler does the conversion for the programmer automatically.*

Another handy function takes a record or array type, and makes a value copy onto the heap, returning a pointer to the copy. The signature is –

```
PROCEDURE BOX(s : T) : POINTER TO T;
```

Where T is a record, array or string type. The function copies the value so that modification of the boxed value does not affect the original. The function is particularly convenient for programs that manipulate character data implemented as dynamically allocated arrays. Thus `BOX("hello")` returns a pointer to array of char of length 6, while `BOX(ptr1^ + ptr2^)` performs a concatenation and allocates a destination array of the required length. If the function is applied to an array of fixed length the return value is an open array of the same length. In the case of character arrays the use of the array stringifier mark "\$" truncates the value at the first nul character, as usual.

Without this function, the construction of a value copy of an open array requires the following tedious construction –

```
VAR a,b : POINTER TO ARRAY OF CHAR;
...
NEW(b, LEN(a));
FOR i := 0 TO LEN(a) DO b[i] := a[i] END;
```

Using the new function, this simply becomes `b := BOX(a^);`

As of version 1.2 a new built-in unsigned byte type has been introduced, for conformance with the .NET CLS. In order to coerce values of signed type to the new type a new function `USHORT`, analogous to the standard `SHORT` function is also introduced. This function has the signature –

```
PROCEDURE USHORT(s : AnyNumericType) : UBYTE;
```

It is a runtime error if the value of the parameter is not within the unsigned byte range.

The final new built-in function, `THROW`, allows programs to throw exceptions, and is described below.

#### 4.8 Deprecated features and warnings

The use of procedure variables and of super-calls are deprecated. Both attract compile-time warning messages. Warnings are also issued in the case of procedures that are not exported, and are not called (or assigned as procedure variables) within their defining module. This situation is usually an error arising from failure to mark the procedure for export.

#### 4.9 Program executable verification

Component Pascal is a type-safe language. Every correct program is type-safe in the same sense that is guaranteed by the .NET virtual object system's verifier. In principle therefore, all output of *gpcp* should be verifiable.

You may test-verify the output of compilation by running the stand-alone program executable verifier `pverify` over the file. Here is an example –

```
C:\CPtree\CPngws> pverify /IL ShowSyms.exe

Microsoft ® COM+ PE Verifier. Version 2000.14.1812.10
Copyright © Microsoft Corp. 1998-2000
```

All Classes and Methods in ShowSyms.exe verified

C:\CPtree\CPngws>

Output might fail to verify if a manually constructed interface to a library does not correspond to the internal metadata of the imported assembly. This potential problem has largely gone away with the use of **N2cps**.

#### 4.10 Unchecked arithmetic

By default, all arithmetic is overflow-checked, and all narrowing assignments are range checked. Sometimes it is necessary to turn off this behaviour. There are two means to do this. One of these is a custom attribute that is applied on a per-procedure basis. Checks may also be turned off from the command line for all compilations in that invocation.

The syntax of the custom attribute is a context sensitive marker that appears immediately after the keyword **BEGIN** in a procedure or module body. The syntax is –

Body → **BEGIN** [ “[ “**UNCHECKED\_ARITHMETIC**” “]” ]  
StatementSequence **END identifier** .

An example of the use of this construct, from the source of the compiler itself, is the identifier hash function –

```
PROCEDURE hashStr(IN str : ARRAY OF CHAR) : INTEGER;
  VAR tot : INTEGER;
      idx : INTEGER;
      len : INTEGER;
BEGIN [UNCHECKED_ARITHMETIC]
  len := LEN(str$);
  tot := 0;
  FOR idx := 0 TO len-1 DO
    INC(tot, tot);
    IF tot < 0 THEN INC(tot) END;
    INC(tot, ORD(str[idx]));
  END;
  RETURN tot MOD size;
END hashStr;
```

This function performs a rotate-and-add computation, in which bits are carried out of the sign bit back into the least significant bit of the variable *tot*. Overflow checking must be turned off, in order to prevent very long identifiers from crashing the compiler.

- ☒ **Important note on parameter passing semantics if you use /target=jvm.**
- ☒ *The semantics of parameter passing on the .NET version are precise. They are also precise in the D-Code version.*
- ☐ *The JVM version of **gpcp** takes liberties with the precise semantics of parameter passing. Actual parameters of unboxed value type that are passed to reference formals are passed by copying. (Unboxed value types are the built-in standard*

*types such as CHAR and INTEGER, together with the pointer types. Structures and arrays are always boxed at runtime in the JVM, and are not affected by this semantic modification.) In the case of formal parameters of **VAR** mode, actual values of unboxed value type are copied **in** and copied **out**. In the case of formal parameters of **OUT** mode the value is copied **out**. This change is necessary in order to obtain reasonable performance on the JVM. This change will not affect the results of your program unless you access the actual of an a reference formal along two paths (either by having two reference formals sharing the same actual, or accessing a static variable directly and through a parameter). You should not write programs that do this! You might also care to know that with this change, the performance of code is good if you have only one such copied parameter, but becomes poor if you have more than one in any frequently called procedure.*

## 5. Exception Handling

Component Pascal does not define exception handling, but it is necessary to deal with foreign libraries that may throw exceptions. There is one new keyword and one new builtin procedure introduced to facilitate this

### 5.1 The RESCUE clause

Procedures, but not modules may include exactly one *RESCUE* clause, at the end of the procedure body. This has syntax –

ProcBody → **BEGIN** Statements [**RESCUE** '(' ident ')' Statements] **END** ident.

The identifier introduced in the parentheses is of type **RTS.NativeException**, and must have a name that is distinct from every other identifier in the local scope.

If any exception is thrown in the body of the procedure, or if any exception is unhandled in a procedure called from this procedure, then the rescue clause is entered with the exception object in the named local variable. This variable is read-only within the rescue clause, and is not known in the rest of the procedure body.

If the program has imported or defined any extensions of the native exception type, filtering may be performed by using the usual type-test syntaxes. The compiler will check that the rescue clause fulfills any contracts implied by the procedure signature. For example, in the case of function procedures the rescue clause must explicitly return a type-correct value, or explicitly throw another exception.

### 5.2 The THROW statement

Code may throw an exception by using the built-in procedure *THROW*. This procedure has two signatures –

```
PROCEDURE THROW(x : RTS.NativeException);
PROCEDURE THROW(x : RTS.NativeString);
```

This may be used anywhere in the program, but is most useful for rethrowing an exception from within a rescue clause.

- ☑ *Remember, if you use these non-standard facilities for exception handling your program source will not be portable to other implementations of Component Pascal. Of course it will still be portable between different implementations of Gardens Point Component Pascal.*

If you want to create an exception object to abort program execution with a meaningful string, the library function **RTS.Throw(msg : ARRAY OF CHAR)** may be used. Exceptions thrown by this library function **can** be caught by a *RESCUE* clause.

## 6. Facilities of the CP Runtime System

### 6.1 Supplied libraries

This release has a small number of libraries supplied. These are –

- **Console**            this library writes strings and numbers to the console
- **Error**             this library writes strings and number to the error stream
- **ProgArgs**        this library provides access to the command line arguments, if any
- **GPText**            a basic library for handling text formatting
- **GPFiles**           defines the supertype of *GPBinFiles.FILE* and *GPTextFiles.FILE*
- **GPBinFiles**      reading and writing binary files
- **GPTextFiles**      reading and writing text files
- **RTS**                access to the facilities of the runtime system

For the most part these libraries are the ones that were required to bootstrap the compiler. More will come later ...

### 6.2 The runtime system RTS.cp

The runtime system provides a variety of low-level access facilities. The source file for this module, **RTS.cp**, is not really the source. This file is a dummy, as is denoted by the context-sensitive mark “*SYSTEM*” appearing before the keyword *MODULE*. All such “modules” are actually implemented in the C# file named *RTS.cs*, and at runtime are found in the assembly *RTS.dll*.

*ProgArgs*, *Console*, and *Error* are also system modules, and have their real source in the same C# file.

Here is the “source” of RTS.

```
(**  These are the user accessible static methods of the CP runtime system.
*    These are the environment-independent ones. Others are in CP*.cp
*    Note: the bodies of these procedures are dummies, this module is
*    compiled with /special. The real code is in RTS.cs or other. *)
```

```
SYSTEM MODULE RTS;
  VAR defaultTarget- : ARRAY 4 OF CHAR;

  TYPE
    CharOpen*        = POINTER TO ARRAY OF CHAR;

  TYPE
    NativeType*      = POINTER TO RECORD END;
    NativeObject*    = POINTER TO RECORD END;
```

```

NativeString*      = POINTER TO RECORD END;
NativeException*   = POINTER TO RECORD END;

PROCEDURE getStr(x : NativeException) : CharOpen;
(** Get error message from Exception *)

PROCEDURE StrToReal*(IN s : ARRAY OF CHAR;
                     OUT r : REAL;
                     OUT ok : BOOLEAN);
(** Parse array into an ieee double REAL *)

PROCEDURE StrToInt*(IN s : ARRAY OF CHAR;
                    OUT i : INTEGER;
                    OUT ok : BOOLEAN);
(** Parse an array into a CP INTEGER *)

PROCEDURE StrToLong*(IN s : ARRAY OF CHAR;
                     OUT i : LONGINT;
                     OUT ok : BOOLEAN);
(** Parse an array into a CP LONGINT *)

PROCEDURE RealToStr*(r : REAL;
                     OUT s : ARRAY OF CHAR);
(** Decode a CP REAL into an array *)

PROCEDURE IntToStr*(i : INTEGER;
                    OUT s : ARRAY OF CHAR);
(** Decode a CP INTEGER into an array *)

PROCEDURE LongToStr*(i : LONGINT;
                     OUT s : ARRAY OF CHAR);
(** Decode a CP INTEGER into an array *)

PROCEDURE realToLongBits*(r : REAL) : LONGINT;
(** Convert ieee double to longint with same bit pattern *)

PROCEDURE longBitsToReal*(l : LONGINT) : REAL;
(** Convert ieee double to a longint with same bit pattern *)

PROCEDURE hiInt*(l : LONGINT) : INTEGER;
(** Get hi-significant word of long integer *)

PROCEDURE loInt*(l : LONGINT) : INTEGER;
(** Get lo-significant word of long integer *)

PROCEDURE Throw*(IN s : ARRAY OF CHAR);
(** Abort execution with an error *)

PROCEDURE GetMillis*() : LONGINT;
(** Get time in milliseconds *)

PROCEDURE GetDateString*(OUT str : ARRAY OF CHAR);
(** Get a date string in some native format *)

PROCEDURE ClassMarker*(o : ANYPTR); (* write class name *)

```



**END RTS.**

The four character *defaultTarget* string will hold “*net*” when running on the .NET platform, and “*jvm*” when running under the Java Runtime Environment.

The word *SYSTEM* in the first line of the definition is a context sensitive mark, rather than a reserved word. This means that the word may be used as an identifier elsewhere in the program. The mark simply indicates that the resources of this module are actually found in the assembly **RTS.dll**. *Console*, *Error* and *ProgArgs* are also *SYSTEM* modules.

## 7. Foreign Language Interface

### 7.1 Accessing the basic underlying types

The underlying types are accessible without any other import other than RTS. At runtime the compiler queries the target flag, or takes the default value if there is no target command option.

If the target is “net” then NativeObject, NativeString and NativeException will be the CLS types **system.Object**, **system.String** and **system.Exception** respectively.

If the target is “jvm” then NativeObject, NativeString and NativeException will be the Java types **java.lang.Object**, **java.lang.String** and **java.lang.Exception** respectively.

In any case, literal strings may be implicitly coerced to either the native string type, or to the native object type. This saves a lot of clutter in code that interfaces to foreign libraries. However, if a CP-style, non-literal string, i.e. a nul-terminated array of char needs to be transformed to a native string, the non-standard built-in function –

**PROCEDURE** MKSTR(**IN** s : **ARRAY OF** CHAR) : RTS.NativeString;  
may be used. See the appendix for an example of using these facilities for working with native string types.

### 7.2 Compiling dummy definition modules

As an interim measure, the compiler has been enhanced so as to allow the construction of metainformation files for foreign language libraries. Such modules must be compiled with the **/special** option.

Foreign language interfaces are denoted by the context sensitive marks **FOREIGN** or **SYSTEM** preceding the keyword **MODULE** at the start of the file. Such “dummy” modules do not contain the code of the foreign language facilities, but simply define the interface to those facilities. Such modules must be compiled with the **/special** option. The system marker has special meaning in the .NET platform, but has the same semantics as foreign in the JVM platform.

When a dummy definition module is compiled there are a small number of syntactic extensions and changes.

- Modules can be given an explicit external name
- Procedures can be given an explicit external name

- Features with protected scope may be defined
- Static features of classes may be defined
- Escaped identifiers may be defined
- Interface types may be defined
- Overloaded names may be given aliases
- Constructors may be given an alias

The syntax - **MODULE Foo["[blah]space"];** declares that this module will be found in assembly "blah" in namespace "space". It is not necessary to use this mechanism if you write the foreign module so that it has the default name as described in Section 3, subsection "*Common Language Specification Names*".

The syntax - **PROCEDURE (x : T)BarII\*["Bar"](i,j : INTEGER);** declares that this procedure has the external name "Bar" and the internal (CP) name "BarII". This mechanism allows overloaded names in the CLS to be given non-overloaded aliases in CP.

The mark "!" is used to declare that a foreign name has protected scope.

If a name clashes with a CP keyword, it should be defined using the back-quote escape.

Here is an example of the syntax that is required to define a foreign interface type.

**TYPE Foo\* = POINTER TO INTERFACE RECORD END;**

The keyword *INTERFACE* is reserved, and such a type cannot declare any fields in the record, nor can it define type-bound procedure which are not declared *ABSTRACT*.

Finally, constructors **must** be declared with the special name ".ctor". Declaring a constructor is not necessary if only the no-arg constructor is required, since **NEW(obj)** works in this case as for other types in CP (see section 8.4 for more detail). If access to constructors with arguments is required, then these are given a CP alias, and are marked as constructors by using the magic explicit name. For the **target=jvm** version, the magic name is "<init>".

### 7.3 Accessing Static Features of Foreign Classes

If a class has been imported from a foreign definition, and the class has static members, these may be accessed by means of a semantic extension to the designator grammar.

Normally, the syntactic construct –

QualifiedIdent {Selector}

is in error if the qualified identifier resolves to a type-identifier. However there are two exceptional cases where this is legal in *gpcp*. If a designator begins –

TypeIdentifier "." Identifier

where

- the type identifier resolves to an imported, foreign type, and
  - the identifier is a static field or constant of the type, or
  - the identifier is a static method of the type
- then this is a legal reference to the static feature of the type.

In order to define such constructs in the syntax of dummy definitions the following syntax is added to the record syntax. Note that these extensions are only valid if the module is compiled with the **/special** command line option.

Record → **RECORD** [ “(“ TypeId “)” { FieldList } [ **STATIC** { StatFeature } ] **END**.  
 StatFeature → ProcHeading | StatConst | StatField .  
 StatConst → identifier “=” ConstExpression .  
 StatField → identifier “:” TypeId .

Procedure headings have the same syntax as elsewhere in the language.

## 8.0 Creating Foreign Definition Modules

This Section is only of relevance if you plan to write your own foreign definition modules. For most users the information in the previous section on the *usage* of these facilities will be sufficient.

### 8.1 Syntax of Foreign Definitions

The syntax of foreign definition is as follows. Unless otherwise defined here, the meanings of non-terminal symbols is the same as in the Component Pascal Report.

GModule → Module | ForeignMod .  
 ForeignMod → ( **FOREIGN** | **SYSTEM** ) **MODULE** *ident* [ *string* ] “;”  
                   ImportList DeclSeq **END** *ident* “.” .  
 DeclSeq → { **CONST** { ConstDecl “;” } | **TYPE** { TypeDecl “:” } | **VAR** { VarDecl “;” } }  
                   { ProcHeading “;” | MethodHeading “;” }  
 ProcHeading → **PROCEDURE** IdentDef [ “[“ *string* “]” ] [FormalPars] .  
 MethodHeading → **PROCEDURE** Receiver IdentDef [ “[“ *string* “]” ] [FormalPars]  
                   [ “;” **NEW** ] [ “;” **ABSTRACT** | **EMPTY** | **EXTENSIBLE** ] .  
 TypeDecl → IdentDef “=” Type .  
 Type → [ **POINTER TO** ] [Attributes] **RECORD** [ “(“ *Qualident* “)” ]  
                   FieldList { “;” FieldList }  
                   [ **STATIC** StaticDecl { “;” StaticDecl } ] **END**  
                   | *other types as in the Report* .  
 StaticDecl → IdentList “;” Type | IdentDef “=” ConstExpr | ProcHeading .  
 Attributes → **ABSTRACT** | **EXTENSIBLE** | **INTERFACE** .

The syntax begins with the context sensitive mark “foreign” or “system”. On the *.NET* platform the system marker indicates that the code will be found in the runtime system assembly. In the *JVM*, where each class file contains a single class, the marker has the same semantic effect as “**FOREIGN**”.

### 8.2 Explicit package or namespace names

The way in which runtime names are generated from module names was described in Section 3.2. In the case of the *JVM* we have the following correspondence.

<i>Component Pascal Name</i>	<i>JVM Name</i>
<b>MODULE</b> ModNm ;	CP.ModNm // package name
<b>TYPE</b> Cls = <b>RECORD</b> ... <b>END</b> ;	CP.ModNm.ModNm_Cls // class name
<b>VAR</b> varNm : Cls ;	CP.ModNm.ModNm.varNm
<b>PROCEDURE</b> ProcNm ( ) ;	CP.ModNm.ModNm.ProcNm ( )

```

PROCEDURE (t : Cls)MthNm();      CP.ModNm.Cls.MthNm()
END ModNm.

```

Notice that in the *JVM* there are no features that are defined outside of classes, so that the static entities *varNm* and *ProcNm* are considered at runtime to belong to an *implicit static class* with the same name as the module name. However, so far as an importing Component Pascal program is concerned, these features will be accessed by the familiar *ModuleName.memberName* syntax.

<i>Component Pascal Name</i>	<i>.NET CLS Name</i>
<b>MODULE</b> ModNm;	[ModNm]ModNm // namespace
<b>TYPE</b> Cls = <b>RECORD</b> ... <b>END</b> ;	[ModNm]ModNm.Cls // class name
<b>VAR</b> varNm : Cls;	[ModNm]ModNm.ModNm::varNm
<b>PROCEDURE</b> ProcNm();	[ModNm]ModNm.ModNm::ProcNm()
<b>PROCEDURE</b> (t : Cls)MthNm();	[ModNm]ModNm.Cls::MthNm()
<b>END</b> ModNm.	

In the virtual object system of *.NET* the situation is similar, with an implicit static class being defined with the same name as the module.

If, as a user, you are writing a foreign definition and plan to implement the library yourself in either Java or in C# (say), then you may define the foreign module in this way and write the foreign code so as to match the default “name mangling” scheme. In this case you may even use the same foreign definition for both versions of *gpcp*, and implement a foreign module on each underlying platform. If on the other hand you are planning to match a foreign definition to an existing library written in Java or C#, then you must override this default naming scheme.

The syntax “**FOREIGN MODULE** *ident* [*string*];” allows an arbitrary package or namespace name to be defined. For example, in order to access the facilities of the package **java.lang.Reflect** a foreign module might begin –

```

FOREIGN MODULE java_lang_Reflect["java.lang.Reflect"];

```

Similarly, in order to access the facilities of the namespace **System.Reflect** in the assembly **mscorlib** a foreign module might begin

```

FOREIGN MODULE mscorlib_System_Reflect
    ["[mscorlib]System.Reflect"];

```

Note that the form of the literal string is different on the two platforms, and thus any such foreign modules will be specific to a particular platform. Notice also that there is no mechanism to explicitly give a name to an implicit static class.

### 8.3 Dealing with overloaded names

Each of the underlying platforms allows name overloading for methods. This feature is deliberately not permitted in Component Pascal. Nevertheless, it is necessary to gain access to library methods that have overloaded names. The option of using explicit external method names facilitates this. Suppose we have two methods, both of which are named *Add()*, one with a single integer parameter, and another with two. We might define these as follows in a foreign definition.

```

PROCEDURE (this : Cls)AddI*["Add"](I : INTEGER),NEW;
PROCEDURE (this : Cls)AddII*["Add"](I,J : INTEGER),NEW;

```

Within the importing CP program the two names are distinct, but the program executable will correctly refer to the underlying overloaded methods. This name-mangling is rather awkward, particularly in the case of parameters of object types.

Since release (1.1) users are able to access the unmangled names of overloaded foreign methods directly. The N2CPS and J2CPS tools create symbol files that have overloaded names, and the compiler will match calls to the intended method. Because this is a language extension, the compiler is strict about matching calls to methods in the presence of automatic coercions. If more than one method matches, when taking into account all legal coercions, *gpcp* will reject the program and require the user to specify the intended coercions of the actual parameters.

#### 8.4 Interfacing to constructors

If a foreign class has a “no-arg” constructor, then this will be implicitly called whenever an object is created by the use of the standard procedure **NEW**. However if it is necessary to access constructors with arguments, then it is possible to define an alias for the constructor in a foreign module. In every case the constructor will be accessed by means of a static, value returning function that returns an object of the constructed class. The fact that this is a constructor must be made known to *gpcp* since the way in which these methods are called differs from other methods. On each underlying platform there is a “magic” name that is used for calling a constructor. On *JVM* the name is “<init>”, while on *.NET* the name is “**.ctor**”. These two strings are used as the explicit string that defines such a procedure in the foreign definition. An example of an interface to a constructor with arguments, in the syntax used by the *Browse* tool, might be –

```
PROCEDURE Init*(width,height : INTEGER) : Rect,CONSTRUCTOR;
```

The identifier “**CONSTRUCTOR**” is not a reserved word, but a context sensitive mark that may be used as an ordinary identifier elsewhere in the program.

Note that this declaration would normally be declared in the static part of the record defining the class “Rect”. Calls to this procedure in a Component Pascal program, such as –

```
rec1 := F.Rect.Init(25,17);
```

would translate into a call to the appropriate one of –

```
namespaceName.Rect::.ctor(int32,int32)
packageName.Rect.<init>(II)
```

Of course, if you extend a foreign class that does not have a public no-arg constructor, then you will not be able to construct values of your own type using **NEW**, since this implicitly calls the no-arg constructor of its super-type. In this case, it is necessary to define a new constructor signature for your extended type. In release 1.2 there are two ways to do this. If the desired constructor has the same signature as the constructor of the supertype, then the first method may be used. In the case of the example above, the required syntax is shown in the following fragment –

```
TYPE MyRect* = POINTER TO RECORD (A.Rect) ... END;
...
PROCEDURE Init*(w,h : INTEGER): MyRect,CONSTRUCTOR;
```

The constructor does not have a body, and simply passes its arguments to the super-type constructor with matching signature.

The new syntax in version 1.2 is considerably more flexible. The Component Pascal definition does not require to have the same signature as the constructor of the super-type. An example of the syntax is –

```
PROCEDURE MkMyRect* (formals) : MyRect , BASE (actuals) ;
    Local-declarations
BEGIN
    Constructor body
    RETURN SELF ;
END MkMyRect ;
```

The identifier “BASE” is not a reserved word, but is a context sensitive mark. It denotes the super-type constructor with the signature matching the types of the actual parameter expressions. This super-type constructor will be called as the first action of the constructor, before the new fields of the derived object are initialized. Within the body of the constructor the object under construction is denoted by the identifier “SELF”. The constructor must return this object along every terminating path of the body. It is an error if the actual parameter expression types in the super-call do not choose a unique super-type constructor.

### 8.5 Declaring static features of classes

Classes in foreign modules may be declared either as records or as pointers to records. However, it is recommended that on the *JVM* platform the pointer form be always used, as a helpful reminder to the user that at runtime the objects will be dynamically allocated. On the *.NET* platform value classes should be declared as plain records, with no explicit base type. On both platforms array types should be declared as pointers to arrays, again reminding the user that all arrays are dynamically (and explicitly) allocated.

In order to access static features of foreign classes, the syntax extension of records given in Section 8.1 must be used. In the optional static section of a record declaration we may define constants, static fields and static (i.e. non type-bound) procedures.

We may consider the following example –

<i>Component Pascal Foreign Definition</i>	<i>Component Pascal Usage</i>
<b>FOREIGN MODULE</b> ModNm;	
<b>TYPE</b> Cls =	ModNm.Cls <i>// class name</i>
<b>POINTER TO RECORD</b>	
<b>STATIC</b>	
statVar* : <b>CHAR</b> ;	ModNm.Cls.statVar
<b>PROCEDURE</b> StatProc() ;	ModNm.Cls.StatProc()
<b>END</b> ;	
<b>END</b> ModNm.	

In this example we select the static member by qualifying the designator by the type-name of the class.

Type-bound methods will be defined lexically outside of the record declaration in the normal CP way, remembering that only the heading is required. On the *.NET* platform the distinction between virtual and instance methods is made automatically. Instance methods are *NEW* but not *EXTENSIBLE*. On the *JVM* platform the possibility of optimizing the calls to such methods are left to the JIT to determine.

☑ *Note that the foreign modules which arise from C# on the .NET platform or are written in Java can never have static features outside of classes. If you are writing the foreign module yourself you may use the default class naming scheme described in Section 3. However if you are matching an existing package, you will need to use the explicit name override described earlier in this Section. This allows you to control the package name, but does not allow you to name an implicit static class for static features. Therefore you will need to use the mechanisms of this sub-section if the package contains any static features.*

## 8.6 Automatic module renaming

Programs written in C# that contain a single class definition only are often created in files that take their name from the name of the class. If you try to match this same structure in Component Pascal, you run into a small difficulty on the *.NET* platform. Suppose you want to export a class “Rename” from a module named “Rename”. In this case the external class name in *.NET* [Rename]Rename.Rename will clash with the name of the synthetic static class. In this case gpcp will automatically rename the static class, by appending two underscores. If the module with the renamed class is imported, gpcp will find the renamed symbol file. In both contexts gpcp will issue a warning that the renaming is taking place –

```
C:\gpcp\work> gpcp Rename.cp UseRename.cp
1 MODULE Rename;
**** -----^ Warning: Default static class has name clash
**** Renaming static class to <__Rename>
#gpcp: <Rename> No errors, and one warning
2 IMPORT Rename, CPmain;
**** -----^ Warning: Looking for an automatically renamed module
**** -----^ Looking for module "Rename" in file <__Rename.cps>
#gpcp: <UseRename> No errors, and one warning
```

## 8. Installing and Trying the Compiler

### 8.1 Installation

The compiler is packaged in a single zip file which is usually unzipped into a directory with a name such as **\gpcp**. There are six sub-directories. These are –

- **bin** the binary files of the compiler
- **docs** the documentation, including this file
- **examples** some example programs
- **libs** contains the simple library files
- **source** the source files
- **work** a working directory to play around with

The bin directory needs to be on your PATH, and the environment variable CPSYM must point to the libs directory. Typical commands are –

```
set CPSYM=. ;C:\gpcp\libs
set PATH=%PATH%;C:\gpcp\bin
```

If you use the installer version (from version 1.1.4) you should not need to do anything other than make responses to the installer’s queries.

## 9. Future Releases

Release 1.1 still has a very limited range of libraries packaged with it, essentially only those needed to bootstrap the compiler. The distribution is sufficient to try out the compiler, and is being updated on a frequent basis. We expect new releases to contain new tools and new libraries.

Updates are announced and available from [www.plasrc.qut.edu.au/ComponentPascal](http://www.plasrc.qut.edu.au/ComponentPascal)

### Changes from 1.1.6

The following changes and corrections are included in the 1.2.0 release.

The semantics of “super-calls” were incorrect in the case that the immediate super-type did not define the method being overridden. In version 1.2 the notation “**Foo**^( )” denotes the overridden method no matter how distant it is in the inheritance hierarchy.

New options have been implemented for output directories, and the default behavior for the /nodebug option is to use the direct PE-file writer. This is significantly faster than going through ilasm. Unfortunately, this new file-writer does not produce debug symbols at this stage. There is separate documentation for the PEAPI component included with this release.

The permitted semantics for constructors with arguments is significantly enhanced. This is of some importance when deriving from types that do not have public no-arg constructors.

### Changes from 1.1.4

The following changes and corrections are included in the 1.1.6 release –

1. Uplevel addressing of reference parameters is now permitted in the .NET release, although this has inexact semantics in some cases.
2. A number of corrections to the JVM code-emitter have been added.
3. The new builtin function BOX has been added.
4. Trapping of types that attempt to indirectly include themselves is improved.
5. An automatic renaming scheme is implemented for modules that attempt to export types with the same name as the module on the .NET platform.

### Changes from 1.1.3

The following changes and corrections are included in the 1.1.4 release –

1. The copyright notice has been revised. GPCP is still open source, but now has a “FreeBSD-like” licence agreement.
2. A correction to the Java class-file emitter now puts correct visibility markers on package-public members. *Appletviewer* didn’t care, but browsers objected!
3. It is now permitted to export type-bound procedures of non-exported types, provided the procedure overrides an exported method of a super-type.



4. More line-markers are emitted to IL. This makes it possible to place a breakpoint on the predicate of a conditional statement, and have the debugger stop on the predicate rather than the next executable statement.
5. The type-resolution code of SymFileRW.cp has been radically revised. It is believed that the code is now immune to certain problems caused by importing foreign libraries with circular dependencies.

## Appendix A – Working with Native Strings

There are some subtleties in converting to native strings. The following example demonstrates several strategies. The example tries to call the *Equals()* method of *System.String* to compare with a Component Pascal literal string.

```
MODULE StringCompare;
IMPORT
  Sys := mscorlib_System,
  CPmain;

VAR type : Sys.Type;
    name : Sys.String;
    ltNm : Sys.String;
    objS : Sys.Object;

BEGIN
  name := type.get_Name();
  ( *
  * This attempt does not work, since literal strings may have several automatic coercions
  * that match different overloads of the Equals() method.
  * )
  IF name.Equals("Blah") THEN END;
  ( *
  * Type assertions only apply to variables. This attempt caused a compiler exception in 1.1.4!
  * )
  IF name.Equals((Sys.String)"Blah") THEN END;
  ( *
  * this next try is invalid syntax, the "cast" construct is a type-assertion, not a conversion
  * )
  IF name.Equals("Blah"(Sys.String)) THEN END;
  ( *
  * Conversions use built-in functions. Here is a non-standard one that
  * converts char-arrays to native strings. This works ...
  * )
  IF name.Equals(MKSTR("Blah")) THEN END;
  ( *
  * In the case of assignments (or non-overloaded method calls), the compiler can work it out
  * by itself without the MKSTR(). Literal char arrays can be assigned to objects or strings.
  * This works.
  * )
  ltNm := "Blah"; ( * gpcp automatically converts the string to System.String * )
  IF name.Equals(ltNm) THEN END;
  ( *
  * In the case of reference variables the type-assertion / cast syntax does work --
  * the following two calls bind to different overloads.
  * )
  objS := "Blah"; ( * gpcp automatically converts the string to System.Object * )
  IF name.Equals(objS) THEN END;
  IF name.Equals(objS(Sys.String)) THEN END;
END StringCompare.
```

Here is the listing output from running gpcp version 1.1.6 on this example-

```

1 MODULE StringCompare;
2   IMPORT
3     Sys := mscorlib_System,
4     CPmain;
5
6   VAR type : Sys.Type;
7       name : Sys.String;
8       ltNm : Sys.String;
9       objS : Sys.Object;
10
11 BEGIN
12   name := type.get_Name();
13   (*
14    * This one does not work, since literal strings
15    * may have several automatic coercions that match
16    * different overloads of the Equals() method.
17    *)
18   IF name.Equals("Blah") THEN END;
19   **** -----^
20   **** Multiple overloaded procedure signatures match this call
21   **** Matches with - Equals(mscorlib_System.String)
22   **** Matches with - Equals(mscorlib_System.Object)
23   (*
24    * Type assertions only apply to variables.
25    * This attempt caused a compiler exception in 1.1.4!
26    *)
27   IF name.Equals((Sys.String)"Blah") THEN END;
28   **** -----^ Not dynamically typed, so you cannot have type-guard
29   **** -----^ Identifier is not a fieldname of the current type
30   **** ')' expected-----^
31   (*
32    * this next try is invalid syntax, the "cast"
33    * construct is a type-assertion, not a conversion
34    *)
35   IF name.Equals("Blah"(Sys.String)) THEN END;
36   **** -----^
37   **** Multiple overloaded procedure signatures match this call
38   **** 'THEN' expected-----^
39   **** Matches with - Equals(mscorlib_System.Object)
40   **** Matches with - Equals(mscorlib_System.String)
41   (*
42    * Conversions use built-in functions. Here is a
43    * non-standard one that converts char-arrays to
44    * native strings. This works ...
45    *)
46   IF name.Equals(MKSTR("Blah")) THEN END;
47   **** -----^
48   **** Warning: Multiple overloaded procedure signatures match this call
49   **** Bound to - Equals(mscorlib_System.String)
50   **** Matches with - Equals(mscorlib_System.Object)
51   (*
52    * In the case of assignments (or non-overloaded
53    * method calls), the compiler can work it out
54    * by itself without the MKSTR(). Literal char
55    * arrays can be assigned to objects or strings.
56    * This works.
57    *)
58   ltNm := "Blah";
59   IF name.Equals(ltNm) THEN END;
60   **** -----^
61   **** Warning: Multiple overloaded procedure signatures match this call
62   **** Bound to - Equals(mscorlib_System.String)
63   **** Matches with - Equals(mscorlib_System.Object)
64   (*
65    * In the case of reference variables, the
66    * type-assertion / cast syntax does work --
67    * the following two calls bind to different
68    * overloads.
69    *)
70   objS := "Blah";
71   IF name.Equals(objS) THEN END;
72   IF name.Equals(objS(Sys.String)) THEN END;
73   **** -----^
74   **** Warning: Multiple overloaded procedure signatures match this call
75   **** Bound to - Equals(mscorlib_System.String)
76   **** Matches with - Equals(mscorlib_System.Object)
77 END StringCompare.

```

## Appendix B – Overriding the Default Naming

The default naming scheme for the .NET version of *gpcp* uses the module name as the stem name for the output files, the CLR assembly name, the namespace name and the dummy static class name. All of these defaults may be overridden as described here. This may be necessary if another component expects a particular naming pattern.

Consider the following short program –

```
MODULE ModId; (* default naming will be used *)  
  TYPE ClsId* = RECORD ... END;  
END ModId.
```

In this case the name of the output file will be `ModId.dll`, the name of the dummy static class will be `[ModId]ModId.ModId`, and the name of the class that represents the record type will be `[ModId]ModId.ClsId`.

It is allowed to follow the module name with a bracketed string that specifies either or both of the assembly name and the namespace name. A typical string would be –

```
MODULE ModId ["[AsmNm]SpcNm"]; (* both *)  
  TYPE ClsId* = RECORD ... END;  
END ModId.
```

In this case the name of the output file will be `AsmNm.dll`, the name of the dummy static class will be `[AsmNm]SpcNm.ModId`, and the name of the class that represents the record type will be `[AsmNm]SpcNm.ClsId`.

In the case that only the assembly name is specified, there is no namespace defined.

```
MODULE ModId ["[AsmNm]"]; (* assembly name only *)  
  TYPE ClsId* = RECORD ... END;  
END ModId.
```

In this case the name of the output file will be `AsmNm.dll`, the name of the dummy static class will be `[AsmNm]ModId`, and the name of the class that represents the record type will be `[AsmNm]ClsId`.

Conversely, if the namespace name is specified, but no assembly name, then the assembly name is taken from the module identifier, as in the default case.

```
MODULE ModId ["SpcNm"]; (* namespace only *)  
  TYPE ClsId* = RECORD ... END;  
END ModId.
```

In this case the name of the output file will be `ModId.dll`, the name of the dummy static class will be `[ModId]SpcNm.ModId`, and the name of the class that represents the record type will be `[ModId]SpcNm.ClsId`.

There is just one special case remaining. In all of the previous cases the name of the dummy static class is taken from the module identifier, with the symbol (metadata) file using the same stem name. If the default name of the static dummy class clashes with the name of an explicit class then the dummy static class will be renamed.

```
MODULE ClsId; (* module name clashes with class id *)  
  TYPE ClsId* = RECORD ... END;  
END ClsId.
```

In this case the name of the output file will be `ClsId.dll`, the name of the dummy static class will be `[ClsId]ClsId.__ClsId`, and the name of the class that represents the record type will be `[ClsId]ClsId.ClsId`. The symbol file will have the name "`__ClsId.cps`" and, as noted earlier, will be automatically found by the compiler if the module name appears in an import list.