

# The CTANGLE processor

(Version 4.2)

	Section	Page
<b>Introduction</b> .....	<a href="#">1</a>	1
Data structures exclusive to CTANGLE .....	<a href="#">19</a>	7
Tokens .....	<a href="#">26</a>	9
<b>Stacks for output</b> .....	<a href="#">31</a>	10
Producing the output .....	<a href="#">41</a>	13
The big output switch .....	<a href="#">48</a>	15
<b>Introduction to the input phase</b> .....	<a href="#">61</a>	20
Inputting the next token .....	<a href="#">68</a>	23
Scanning a macro definition .....	<a href="#">82</a>	30
Scanning a section .....	<a href="#">90</a>	34
<b>Index</b> .....	<a href="#">104</a>	38

Copyright © 1987, 1990, 1993, 2000 Silvio Levy and Donald E. Knuth

Permission is granted to make and distribute verbatim copies of this document provided that the copyright notice and this permission notice are preserved on all copies.

Permission is granted to copy and distribute modified versions of this document under the conditions for verbatim copying, provided that the entire resulting derived work is given a different name and distributed under the terms of a permission notice identical to this one.

**1. Introduction.** This is the CTANGLE program by Silvio Levy and Donald E. Knuth, based on TANGLE by Knuth. We are thankful to Nelson Beebe, Hans-Hermann Bode (to whom the C++ adaptation is due), Klaus Guntermann, Norman Ramsey, Tomas Rokicki, Joachim Schnitter, Joachim Schrod, Lee Wittenberg, and others who have contributed improvements.

The “banner line” defined here should be changed whenever CTANGLE is modified.

```
#define banner "This is CTANGLE (Version 4.2)"
  < Include files 4 >
  < Preprocessor definitions >
  < Common code for CWEAVE and CTANGLE 3 >
  < Typedef declarations 19 >
  < Private variables 20 >
  < Predeclaration of procedures 8 >
```

**2.** CTANGLE has a fairly straightforward outline. It operates in two phases: First it reads the source file, saving the C code in compressed form; then it shuffles and outputs the code.

Please read the documentation for `common`, the set of routines common to CTANGLE and CWEAVE, before proceeding further.

```
int main(int ac, char **av)
{
  argc ← ac; argv ← av; program ← ctangle; < Set initial values 21 >
  common_init();
  if (show_banner) puts(banner);    ▷ print a “banner line” <
  phase_one();    ▷ read all the user’s text and compress it into tok_mem <
  phase_two();    ▷ output the contents of the compressed tables <
  return wrap_up();    ▷ and exit gracefully <
}
```

**3.** The next few sections contain stuff from the file “`common.w`” that must be included in both “`ctangle.w`” and “`cweave.w`”. It appears in file “`common.h`”, which is also included in “`common.w`” to propagate possible changes from this COMMON interface consistently.

First comes general stuff:

```
#define ctangle false
#define cweave true
< Common code for CWEAVE and CTANGLE 3 > ≡
typedef bool boolean;
typedef uint8_t eight_bits;
typedef uint16_t sixteen_bits;
extern boolean program;    ▷ CWEAVE or CTANGLE? <
extern int phase;    ▷ which phase are we in? <
```

See also sections 5, 6, 7, 9, 10, 12, 14, and 15.

This code is used in section 1.

## 4. Interface to the standard C library:

```

⟨Include files 4⟩ ≡
#include <ctype.h>      ▷ definition of isalpha, isdigit and so on ◁
#include <stdbool.h>    ▷ definition of bool, true and false ◁
#include <stddef.h>     ▷ definition of ptrdiff_t ◁
#include <stdint.h>     ▷ definition of uint8_t and uint16_t ◁
#include <stdlib.h>     ▷ definition of getenv and exit ◁
#include <stdio.h>      ▷ definition of printf and friends ◁
#include <string.h>     ▷ definition of strlen, strcmp and so on ◁

This code is used in section 1.

```

## 5. Code related to the character set:

```

#define and_and  °4      ▷ '&&'; corresponds to MIT's ∧ ◁
#define lt_lt   °20     ▷ '<<'; corresponds to MIT's ⊂ ◁
#define gt_gt   °21     ▷ '>>'; corresponds to MIT's ⊃ ◁
#define plus_plus °13    ▷ '++'; corresponds to MIT's ↑ ◁
#define minus_minus °1   ▷ '--'; corresponds to MIT's ↓ ◁
#define minus_gt °31     ▷ '->'; corresponds to MIT's → ◁
#define non_eq °32      ▷ '!='; corresponds to MIT's ≠ ◁
#define lt_eq °34       ▷ '<='; corresponds to MIT's ≤ ◁
#define gt_eq °35       ▷ '>='; corresponds to MIT's ≥ ◁
#define eq_eq °36       ▷ '=='; corresponds to MIT's ≡ ◁
#define or_or °37       ▷ '||'; corresponds to MIT's ∨ ◁
#define dot_dot_dot °16  ▷ '...'; corresponds to MIT's ∞ ◁
#define colon_colon °6   ▷ '::'; corresponds to MIT's ∈ ◁
#define period_ast °26   ▷ '.*'; corresponds to MIT's ⊗ ◁
#define minus_gt_ast °27 ▷ '->*'; corresponds to MIT's ↗ ◁

```

```

⟨Common code for CWEAVE and CTANGLE 3⟩ +≡
extern char section_text[];    ▷ text being sought for ◁
extern char *section_text_end; ▷ end of section_text ◁
extern char *id_first;        ▷ where the current identifier begins in the buffer ◁
extern char *id_loc;          ▷ just after the current identifier in the buffer ◁

```

## 6. Code related to input routines:

```

#define xisalpha(c) (isalpha((eight_bits) c) ∧ ((eight_bits) c < °200))
#define xisdigit(c) (isdigit((eight_bits) c) ∧ ((eight_bits) c < °200))
#define xisspace(c) (isspace((eight_bits) c) ∧ ((eight_bits) c < °200))
#define xislower(c) (islower((eight_bits) c) ∧ ((eight_bits) c < °200))
#define xisupper(c) (isupper((eight_bits) c) ∧ ((eight_bits) c < °200))
#define xisxdigit(c) (isxdigit((eight_bits) c) ∧ ((eight_bits) c < °200))

```

```

⟨Common code for CWEAVE and CTANGLE 3⟩ +≡
extern char buffer[];    ▷ where each line of input goes ◁
extern char *buffer_end; ▷ end of buffer ◁
extern char *loc;        ▷ points to the next character to be read from the buffer ◁
extern char *limit;      ▷ points to the last character in the buffer ◁

```

## 7. Code related to file handling:

```

format line x    ▷ make line an unreserved word <
#define max_include_depth 10
    ▷ maximum number of source files open simultaneously, not counting the change file <
#define max_file_name_length 1024
#define cur_file file[include_depth]    ▷ current file <
#define cur_file_name file_name[include_depth]    ▷ current file name <
#define cur_line line[include_depth]    ▷ number of current line in current file <
#define web_file file[0]    ▷ main source file <
#define web_file_name file_name[0]    ▷ main source file name <
{Common code for CWEAVE and CTANGLE 3} +=
extern int include_depth;    ▷ current level of nesting <
extern FILE *file[];    ▷ stack of non-change files <
extern FILE *change_file;    ▷ change file <
extern char file_name[][][max_file_name_length];    ▷ stack of non-change file names <
extern char change_file_name[];    ▷ name of change file <
extern int line[];    ▷ number of current line in the stacked files <
extern int change_line;    ▷ number of current line in change file <
extern int change_depth;    ▷ where @y originated during a change <
extern boolean input_has_ended;    ▷ if there is no more input <
extern boolean changing;    ▷ if the current line is from change_file <
extern boolean web_file_open;    ▷ if the web file is being read <

```

## 8. {Predeclaration of procedures 8} =

```

extern boolean get_line(void);    ▷ inputs the next line <
extern void check_complete(void);    ▷ checks that all changes were picked up <
extern void reset_input(void);    ▷ initialize to read the web file and change file <

```

See also sections 11, 13, 16, 30, 35, 39, 44, 49, 53, 65, 70, 84, 91, 99, and 101.

This code is used in section 1.

## 9. Code related to section numbers:

```

{Common code for CWEAVE and CTANGLE 3} +=
extern sixteen_bits section_count;    ▷ the current section number <
extern boolean changed_section[];    ▷ is the section changed? <
extern boolean change_pending;    ▷ is a decision about change still unclear? <
extern boolean print_where;    ▷ tells CTANGLE to print line and file info <

```

10. Code related to identifier and section name storage:

```
#define length(c) (size_t)((c+1)-byte_start - (c)-byte_start)    ▷ the length of a name ◁
#define print_id(c) term_write((c)-byte_start, length((c)))      ▷ print identifier ◁
#define llink link        ▷ left link in binary search tree for section names ◁
#define rlink dummy.Rlink  ▷ right link in binary search tree for section names ◁
#define root name_dir->rlink  ▷ the root of the binary search tree for section names ◁
```

⟨Common code for CWEAVE and CTANGLE 3⟩ +≡

```
typedef struct name_info {
    char *byte_start;    ▷ beginning of the name in byte_mem ◁
    struct name_info *link;
    union {
        struct name_info *Rlink;    ▷ right link in binary search tree for section names ◁
        char Ilk;    ▷ used by identifiers in CWEAVE only ◁
    } dummy;
    void *equiv_or_xref;    ▷ info corresponding to names ◁
} name_info;    ▷ contains information about an identifier or section name ◁
typedef name_info *name_pointer;    ▷ pointer into array of name_infos ◁
typedef name_pointer *hash_pointer;
extern char byte_mem[];    ▷ characters of names ◁
extern char *byte_mem_end;    ▷ end of byte_mem ◁
extern char *byte_ptr;    ▷ first unused position in byte_mem ◁
extern name_info name_dir[];    ▷ information about names ◁
extern name_pointer name_dir_end;    ▷ end of name_dir ◁
extern name_pointer name_ptr;    ▷ first unused position in name_dir ◁
extern name_pointer hash[];    ▷ heads of hash lists ◁
extern hash_pointer hash_end;    ▷ end of hash ◁
extern hash_pointer h;    ▷ index into hash-head array ◁
```

11. ⟨Predeclaration of procedures 8⟩ +≡

```
extern boolean names_match(name_pointer, const char *, size_t, eight_bits);
extern name_pointer id_lookup(const char *, const char *, char);
    ▷ looks up a string in the identifier table ◁
extern name_pointer section_lookup(char *, char *, boolean);    ▷ finds section name ◁
extern void init_node(name_pointer);
extern void init_p(name_pointer, eight_bits);
extern void print_prefix_name(name_pointer);
extern void print_section_name(name_pointer);
extern void sprint_section_name(char *, name_pointer);
```

## 12. Code related to error handling:

```

#define spotless 0    ▷ history value for normal jobs ◁
#define harmless_message 1    ▷ history value when non-serious info was printed ◁
#define error_message 2    ▷ history value when an error was noted ◁
#define fatal_message 3    ▷ history value when we had to stop prematurely ◁
#define mark_harmless
{
    if (history ≡ spotless) history ← harmless_message;
}
#define mark_error history ← error_message
#define confusion(s) fatal("!_This_can't_happen:_", s)
⟨Common code for CWEAVE and CTANGLE 3⟩ +≡
extern int history;    ▷ indicates how bad this run was ◁

```

## 13. ⟨Predeclaration of procedures 8⟩ +≡

```

extern int wrap_up(void);    ▷ indicate history and exit ◁
extern void err_print(const char *);    ▷ print error message and context ◁
extern void fatal(const char *, const char *);    ▷ issue error message and die ◁
extern void overflow(const char *);    ▷ succumb because a table has overflowed ◁

```

## 14. Code related to command line arguments:

```

#define show_banner flags['b']    ▷ should the banner line be printed? ◁
#define show_progress flags['p']    ▷ should progress reports be printed? ◁
#define show_stats flags['s']    ▷ should statistics be printed at end of run? ◁
#define show_happiness flags['h']    ▷ should lack of errors be announced? ◁
#define make_xrefs flags['x']    ▷ should cross references be output? ◁
⟨Common code for CWEAVE and CTANGLE 3⟩ +≡
extern int argc;    ▷ copy of ac parameter to main ◁
extern char **argv;    ▷ copy of av parameter to main ◁
extern char C_file_name[];    ▷ name of C_file ◁
extern char tex_file_name[];    ▷ name of tex_file ◁
extern char idx_file_name[];    ▷ name of idx_file ◁
extern char scn_file_name[];    ▷ name of scn_file ◁
extern boolean flags[];    ▷ an option for each 7-bit code ◁

```

## 15. Code related to output:

```

#define update_terminal fflush(stdout)    ▷ empty the terminal output buffer ◁
#define new_line putchar('n')
#define putxchar putchar
#define term_write(a, b) fflush(stdout), fwrite(a, sizeof(char), b, stdout)
#define C_printf(c, a) fprintf(C_file, c, a)
#define C_putc(c) putc(c, C_file)    ▷ isn't C wonderfully consistent? ◁
⟨Common code for CWEAVE and CTANGLE 3⟩ +≡
extern FILE *C_file;    ▷ where output of CTANGLE goes ◁
extern FILE *tex_file;    ▷ where output of CWEAVE goes ◁
extern FILE *idx_file;    ▷ where index from CWEAVE goes ◁
extern FILE *scn_file;    ▷ where list of sections from CWEAVE goes ◁
extern FILE *active_file;    ▷ currently active file for CWEAVE output ◁

```

16. The procedure that gets everything rolling:

```
<Predeclaration of procedures 8> +≡
  extern void common_init(void);
  extern void print_stats(void);
```

17. The following parameters were sufficient in the original WEB to handle T<sub>E</sub>X, so they should be sufficient for most applications of CWEB.

```
#define max_bytes 1000000    ▷ the number of bytes in identifiers, index entries, and section names <
#define max_toks 1000000    ▷ number of bytes in compressed C code <
#define max_names 10239     ▷ number of identifiers, strings, section names; must be less than 10240 <
#define max_sections 4000    ▷ greater than the total number of sections <
#define max_texts 10239     ▷ number of replacement texts, must be less than 10240 <
#define longest_name 10000    ▷ file and section names and section texts shouldn't be longer than this <
#define stack_size 500       ▷ number of simultaneous levels of macro expansion <
#define buf_size 1000        ▷ maximum length of input line, plus one <
#define long_buf_size (buf_size + longest_name)    ▷ for CWEAVE <
```

18. End of COMMON interface.

**19. Data structures exclusive to CTANGLE.** We’ve already seen that the *byte\_mem* array holds the names of identifiers, strings, and sections; the *tok\_mem* array holds the replacement texts for sections. Allocation is sequential, since things are deleted only during Phase II, and only in a last-in-first-out manner.

A **text** variable is a structure containing a pointer into *tok\_mem*, which tells where the corresponding text starts, and an integer *text.link*, which, as we shall see later, is used to connect pieces of text that have the same name. All the **texts** are stored in the array *text\_info*, and we use a **text\_pointer** variable to refer to them.

The first position of *tok\_mem* that is unoccupied by replacement text is called *tok\_ptr*, and the first unused location of *text\_info* is called *text\_ptr*. Thus we usually have the identity *text\_ptr*→*tok\_start*  $\equiv$  *tok\_ptr*.

⟨ Typedef declarations 19 ⟩  $\equiv$

```
typedef struct {
    eight_bits *tok_start;    ▷ pointer into tok_mem ◁
    sixteen_bits text_link;   ▷ relates replacement texts ◁
} text;
typedef text *text_pointer;
```

See also section 31.

This code is used in section 1.

**20.** ⟨ Private variables 20 ⟩  $\equiv$

```
static text text_info[max_texts];
static text_pointer text_info_end  $\leftarrow$  text_info + max_texts - 1;
static text_pointer text_ptr;    ▷ first unused position in text_info ◁
static eight_bits tok_mem[max_toks];
static eight_bits *tok_mem_end  $\leftarrow$  tok_mem + max_toks - 1;
static eight_bits *tok_ptr;    ▷ first unused position in tok_mem ◁
```

See also sections 26, 32, 37, 42, 45, 52, 57, 62, 66, 68, and 82.

This code is used in section 1.

**21.** ⟨ Set initial values 21 ⟩  $\equiv$

```
text_info→tok_start  $\leftarrow$  tok_ptr  $\leftarrow$  tok_mem; text_ptr  $\leftarrow$  text_info + 1; text_ptr→tok_start  $\leftarrow$  tok_mem;
▷ this makes replacement text 0 of length zero ◁
```

See also sections 23, 27, 46, 58, 63, and 78.

This code is used in section 2.

**22.** If *p* is a pointer to a section name, *p-equiv* is a pointer to its replacement text, an element of the array *text\_info*.

```
#define equiv equiv_or_xref    ▷ info corresponding to names ◁
```

**23.** ⟨ Set initial values 21 ⟩  $+\equiv$

```
init_node(name_dir);    ▷ the undefined section has no replacement text ◁
```

**24.** Here’s the procedure that decides whether a name of length *l* starting at position *first* equals the identifier pointed to by *p*:

```
boolean names_match(name_pointer p,    ▷ points to the proposed match ◁
    const char *first,    ▷ position of first character of string ◁
    size_t l,    ▷ length of identifier ◁
    eight_bits t)    ▷ not used by TANGLE ◁
{ (void) t;
  if (length(p)  $\neq$  l) return false;
  return  $\neg$ strcmp(first, p→byte_start, l);
}
```



**25.** The common lookup routine refers to separate routines *init\_node* and *init\_p* when the data structure grows. Actually *init\_p* is called only by **CWEAVE**, but we need to declare a dummy version so that the loader won't complain of its absence.

```
void init_node(name_pointer node)
{
    node→equiv ← (void *) text_info;
}

void init_p(name_pointer p, eight_bits t)
{ (void) p; (void) t; }
```

**26. Tokens.** Replacement texts, which represent C code in a compressed format, appear in *tok\_mem* as mentioned above. The codes in these texts are called ‘tokens’; some tokens occupy two consecutive eight-bit byte positions, and the others take just one byte.

If  $p$  points to a replacement text,  $p\text{-tok\_start}$  is the *tok\_mem* position of the first eight-bit code of that text. If  $p\text{-text\_link} \equiv 0$ , this is the replacement text for a macro, otherwise it is the replacement text for a section. In the latter case  $p\text{-text\_link}$  is either equal to *section\_flag*, which means that there is no further text for this section, or  $p\text{-text\_link}$  points to a continuation of this replacement text; such links are created when several sections have C texts with the same name, and they also tie together all the C texts of unnamed sections. The replacement text pointer for the first unnamed section appears in *text\_info-text\_link*, and the most recent such pointer is *last\_unnamed*.

```
#define section_flag max_texts    ▷ final text_link in section replacement texts ◁
⟨Private variables 20⟩ +=
    static text_pointer last_unnamed;    ▷ most recent replacement text of unnamed section ◁
```

**27.** ⟨Set initial values 21⟩ +=  
 $last\_unnamed \leftarrow text\_info; \text{ text\_info-text\_link} \leftarrow macro;$

**28.** If the first byte of a token is less than  $^{\circ}200$ , the token occupies a single byte. Otherwise we make a sixteen-bit token by combining two consecutive bytes  $a$  and  $b$ . If  $^{\circ}200 \leq a < ^{\circ}250$ , then  $(a - ^{\circ}200) \times 2^8 + b$  points to an identifier; if  $^{\circ}250 \leq a < ^{\circ}320$ , then  $(a - ^{\circ}250) \times 2^8 + b$  points to a section name (or, if it has the special value *output\_defs\_flag*, to the area where the preprocessor definitions are stored); and if  $^{\circ}320 \leq a < ^{\circ}400$ , then  $(a - ^{\circ}320) \times 2^8 + b$  is the number of the section in which the current replacement text appears.

Codes less than  $^{\circ}200$  are 7-bit **char** codes that represent themselves. Some of the 7-bit codes will not be present, however, so we can use them for special purposes. The following symbolic names are used:

*join* denotes the concatenation of adjacent items with no space or line breaks allowed between them (the @& operation of CWEB).

*string* denotes the beginning or end of a string, verbatim construction or numerical constant.

```
#define string ^2    ▷ takes the place of extended ASCII α ◁
#define join ^177    ▷ takes the place of ASCII delete ◁
#define output_defs_flag (2 * ^24000 - 1)
```

**29.** The following procedure is used to enter a two-byte value into *tok\_mem* when a replacement text is being generated.

```
static void store_two_bytes(sixteen_bits x)
{
    if (tok_ptr + 2 > tok_mem_end) overflow("token");
    *tok_ptr++ ← x >> 8;    ▷ store high byte ◁
    *tok_ptr++ ← x & ^377;    ▷ store low byte ◁
}
```

**30.** ⟨Predeclaration of procedures 8⟩ += **static void store\_two\_bytes(sixteen\_bits);**

**31. Stacks for output.** The output process uses a stack to keep track of what is going on at different “levels” as the sections are being written out. Entries on this stack have five parts:

*end\_field* is the *tok\_mem* location where the replacement text of a particular level will end;  
*byte\_field* is the *tok\_mem* location from which the next token on a particular level will be read;  
*name\_field* points to the name corresponding to a particular level;  
*repl\_field* points to the replacement text currently being read at a particular level;  
*section\_field* is the section number, or zero if this is a macro.

The current values of these five quantities are referred to quite frequently, so they are stored in a separate place instead of in the *stack* array. We call the current values *cur\_end*, *cur\_byte*, *cur\_name*, *cur\_repl*, and *cur\_section*.

The global variable *stack\_ptr* tells how many levels of output are currently in progress. The end of all output occurs when the stack is empty, i.e., when *stack\_ptr*  $\equiv$  *stack*.

(Typedef declarations 19)  $\equiv$

```
typedef struct {
    eight_bits *end_field;    ▷ ending location of replacement text ◁
    eight_bits *byte_field;   ▷ present location within replacement text ◁
    name_pointer name_field;   ▷ byte_start index for text being output ◁
    text_pointer repl_field;    ▷ tok_start index for text being output ◁
    sixteen_bits section_field; ▷ section number or zero if not a section ◁
} output_state;
typedef output_state *stack_pointer;
```

**32. #define** *cur\_end* *cur\_state.end\_field* ▷ current ending location in *tok\_mem* ◁  
**#define** *cur\_byte* *cur\_state.byte\_field* ▷ location of next output byte in *tok\_mem* ◁  
**#define** *cur\_name* *cur\_state.name\_field* ▷ pointer to current name being expanded ◁  
**#define** *cur\_repl* *cur\_state.repl\_field* ▷ pointer to current replacement text ◁  
**#define** *cur\_section* *cur\_state.section\_field* ▷ current section number being expanded ◁

(Private variables 20)  $\equiv$

```
static output_state cur_state;    ▷ cur_end, cur_byte, cur_name, cur_repl, and cur_section ◁
static output_state stack[stack_size + 1];    ▷ info for non-current levels ◁
static stack_pointer stack_end  $\leftarrow$  stack + stack_size;    ▷ end of stack ◁
static stack_pointer stack_ptr;    ▷ first unused location in the output state stack ◁
```

**33.** To get the output process started, we will perform the following initialization steps. We may assume that *text\_info-text\_link* is nonzero, since it points to the C text in the first unnamed section that generates code; if there are no such sections, there is nothing to output, and an error message will have been generated before we do any of the initialization.

(Initialize the output stacks 33)  $\equiv$

```
stack_ptr  $\leftarrow$  stack + 1; cur_name  $\leftarrow$  name_dir; cur_repl  $\leftarrow$  text_info-text_link + text_info;
cur_byte  $\leftarrow$  cur_repl-tok_start; cur_end  $\leftarrow$  (cur_repl + 1)-tok_start; cur_section  $\leftarrow$  0;
```

This code is used in section 48.

34. When the replacement text for name  $p$  is to be inserted into the output, the following subroutine is called to save the old level of output and get the new one going.

We assume that the C compiler can copy structures.

```
static void push_level(    ▷ suspends the current level ◁
    name_pointer p)
{
    if (stack_ptr == stack_end) overflow("stack");
    *stack_ptr ← cur_state; stack_ptr++;
    if (p ≠ Λ) {    ▷ p ≡ Λ means we are in output_defs ◁
        cur_name ← p; cur_repl ← (text_pointer) p-equiv; cur_byte ← cur_repl-tok_start;
        cur_end ← (cur_repl + 1)-tok_start; cur_section ← 0;
    }
}
```

35. ⟨Predeclaration of procedures 8⟩ +≡

```
static void push_level(name_pointer);
static void pop_level(boolean);
```

36. When we come to the end of a replacement text, the *pop\_level* subroutine does the right thing: It either moves to the continuation of this replacement text or returns the state to the most recently stacked level.

```
static void pop_level(    ▷ do this when cur_byte reaches cur_end ◁
    boolean flag)    ▷ flag ≡ false means we are in output_defs ◁
{
    if (flag ∧ cur_repl-text_link < section_flag) {    ▷ link to a continuation ◁
        cur_repl ← cur_repl-text_link + text_info;    ▷ stay on the same level ◁
        cur_byte ← cur_repl-tok_start; cur_end ← (cur_repl + 1)-tok_start; return;
    }
    stack_ptr--;    ▷ go down to the previous level ◁
    if (stack_ptr > stack) cur_state ← *stack_ptr;
}
```

37. The heart of the output procedure is the function *get\_output*, which produces the next token of output and sends it on to the lower-level function *out\_char*. The main purpose of *get\_output* is to handle the necessary stacking and unstacking. It sends the value *section\_number* if the next output begins or ends the replacement text of some section, in which case *cur\_val* is that section's number (if beginning) or the negative of that value (if ending). (A section number of 0 indicates not the beginning or ending of a section, but a *#line* command.) And it sends the value *identifier* if the next output is an identifier, in which case *cur\_val* points to that identifier name.

```
#define section_number °201    ▷ code returned by get_output for section numbers ◁
#define identifier °202    ▷ code returned by get_output for identifiers ◁
```

⟨Private variables 20⟩ +≡

```
static int cur_val;    ▷ additional information corresponding to output token ◁
```

38. If *get\_output* finds that no more output remains, it returns with *stack\_ptr*  $\equiv$  *stack*.

```
static void get_output(void)    ▷ sends next token to out_char ◁
{
  sixteen_bits a;             ▷ value of current byte ◁

restart:
  if (stack_ptr  $\equiv$  stack) return;
  if (cur_byte  $\equiv$  cur_end) {
    cur_val  $\leftarrow$  -((int) cur_section);    ▷ cast needed because of sign extension ◁
    pop_level(true);
    if (cur_val  $\equiv$  0) goto restart;
    out_char(section_number); return;
  }
  a  $\leftarrow$  *cur_byte++;
  if (out_state  $\equiv$  verbatim  $\wedge$  a  $\neq$  string  $\wedge$  a  $\neq$  constant  $\wedge$  a  $\neq$  '\n') C_putc(a);
    ▷ a high-bit character can occur in a string ◁
  else if (a < °200) out_char(a);    ▷ one-byte token ◁
  else {
    a  $\leftarrow$  (a - °200) * °400 + *cur_byte++;
    switch (a/°24000) {    ▷ °24000  $\equiv$  (°250 - °200) * °400 ◁
    case 0: cur_val  $\leftarrow$  a; out_char(identifier); break;
    case 1:
      if (a  $\equiv$  output_defs_flag) output_defs();
      else ◁ Expand section a - °24000, goto restart 40 ◁
      break;
    default: cur_val  $\leftarrow$  a - °50000;
      if (cur_val > 0) cur_section  $\leftarrow$  cur_val;
      out_char(section_number);
    }
  }
}
```

39. ◁ Predeclaration of procedures 8 ◁  $\equiv$  static void get\_output(void);

40. The user may have forgotten to give any C text for a section name, or the C text may have been associated with a different name by mistake.

```
◁ Expand section a - °24000, goto restart 40 ◁  $\equiv$ 
{
  a -= °24000;
  if ((a + name_dir)-equiv  $\neq$  (void *) text_info) push_level(a + name_dir);
  else if (a  $\neq$  0) {
    fputs("\n!_Not_present:_<", stdout); print_section_name(a + name_dir); err_print(">");
  }
  goto restart;
}
```

This code is used in section 38.

**41. Producing the output.** The *get\_output* routine above handles most of the complexity of output generation, but there are two further considerations that have a nontrivial effect on CTANGLE's algorithms.

**42.** First, we want to make sure that the output has spaces and line breaks in the right places (e.g., not in the middle of a string or a constant or an identifier, not at a '@&' position where quantities are being joined together, and certainly after an = because the C compiler thinks == is ambiguous).

The output process can be in one of following states:

*num\_or\_id* means that the last item in the buffer is a number or identifier, hence a blank space or line break must be inserted if the next item is also a number or identifier.

*unbreakable* means that the last item in the buffer was followed by the '@&' operation that inhibits spaces between it and the next item.

*verbatim* means we're copying only character tokens, and that they are to be output exactly as stored.

This is the case during strings, verbatim constructions and numerical constants.

*post\_slash* means we've just output a slash.

*normal* means none of the above.

Furthermore, if the variable *protect* is *true*, newlines are preceded by a '\'.

```
#define normal 0      ▷ non-unusual state <
#define num_or_id 1    ▷ state associated with numbers and identifiers <
#define post_slash 2    ▷ state following a / <
#define unbreakable 3  ▷ state associated with @& <
#define verbatim 4     ▷ state in the middle of a string <
<Private variables 20> +=
static eight_bits out_state;    ▷ current status of partial output <
static boolean protect;        ▷ should newline characters be quoted? <
```

**43.** Here is a routine that is invoked when we want to output the current line. During the output process, *cur\_line* equals the number of the next line to be output.

```
static void flush_buffer(void)    ▷ writes one line to output file <
{
    C_putc('\n');
    if (cur_line % 100 == 0 & show_progress) {
        printf(".");
        if (cur_line % 500 == 0) printf("%d", cur_line);
        update_terminal;    ▷ progress report <
    }
    cur_line++;
}
```

**44.** <Predeclaration of procedures 8> += **static void flush\_buffer(void);**

**45.** Second, we have modified the original TANGLE so that it will write output on multiple files. If a section name is introduced in at least one place by @(< instead of @<, we treat it as the name of a file. All these special sections are saved on a stack, *output\_files*. We write them out after we've done the unnamed section.

```
#define max_files 256
<Private variables 20> +=
static name_pointer output_files[max_files];
static name_pointer *cur_out_file, *end_output_files, *an_output_file;
static char cur_section_name_char;    ▷ is it '<' or '(' <
static char output_file_name[longest_name + 1];    ▷ name of the file <
```

46. We make *end\_output\_files* point just beyond the end of *output\_files*. The stack pointer *cur\_out\_file* starts out there. Every time we see a new file, we decrement *cur\_out\_file* and then write it in.

⟨Set initial values 21⟩ +=

```
cur_out_file ← end_output_files ← output_files + max_files;
```

47. ⟨If it's not there, add *cur\_section\_name* to the output file stack, or complain we're out of room 47⟩ =

```
{
  for (an_output_file ← cur_out_file; an_output_file < end_output_files; an_output_file++)
    if (*an_output_file ≡ cur_section_name) break;
  if (an_output_file ≡ end_output_files) {
    if (cur_out_file > output_files) *--cur_out_file ← cur_section_name;
    else {
      overflow("output_files");
    }
  }
}
```

This code is used in section 77.

**48. The big output switch.** Here then is the routine that does the output.

```
static void phase_two(void)
{
    web_file_open ← false; cur_line ← 1; ⟨Initialize the output stacks 33⟩
    ⟨Output macro definitions if appropriate 51⟩
    if (text_info→text_link ≡ macro ∧ cur_out_file ≡ end_output_files) {
        fputs("\n!_No_program_text_was_specified.", stdout); mark_harmless;
    }
    else {
        if (cur_out_file ≡ end_output_files) {
            if (show_progress) printf("\nWriting_the_output_file_(%s):", C_file_name);
        }
        else {
            if (show_progress) {
                fputs("\nWriting_the_output_files:", stdout); printf("_(%s)", C_file_name);
                update_terminal;
            }
            if (text_info→text_link ≡ macro) goto writeloop;
        }
        while (stack_ptr > stack) get_output();
        flush_buffer();
    writeloop: ⟨Write all the named output files 50⟩
        if (show_happiness) {
            if (show_progress) new_line;
            fputs("Done.", stdout);
        }
    }
}
```

**49.** ⟨Predeclaration of procedures 8⟩ +≡ **static void phase\_two(void);**

**50.** To write the named output files, we proceed as for the unnamed section. The only subtlety is that we have to open each one.

⟨Write all the named output files 50⟩ ≡

```
for (an_output_file ← end_output_files; an_output_file > cur_out_file; ) {
    an_output_file--; sprint_section_name(output_file_name, *an_output_file); fclose(C_file);
    if ((C_file ← fopen(output_file_name, "wb")) ≡ Λ)
        fatal(!_Can_open_output_file, output_file_name);
    if (show_progress) {
        printf("\n(%s)", output_file_name); update_terminal;
    }
    cur_line ← 1; stack_ptr ← stack + 1; cur_name ← (*an_output_file); cur_repl ← (text_pointer)
        cur_name→equiv; cur_byte ← cur_repl→tok_start; cur_end ← (cur_repl + 1)→tok_start;
    while (stack_ptr > stack) get_output();
    flush_buffer();
}
```

This code is used in section 48.



51. If a `@h` was not encountered in the input, we go through the list of replacement texts and copy the ones that refer to macros, preceded by the `#define` preprocessor command.

⟨Output macro definitions if appropriate 51⟩  $\equiv$

```
if ( $\neg$ output_defs_seen) output_defs();
```

This code is used in section 48.

52. ⟨Private variables 20⟩  $+\equiv$

```
static boolean output_defs_seen  $\leftarrow$  false;
```

53. ⟨Predeclaration of procedures 8⟩  $+\equiv$

```
static void output_defs(void);
```

```
static void out_char(eight_bits);
```

54. `static void output_defs(void)`

```
{
  sixteen_bits a;
  push_level( $\Lambda$ );
  for (cur_text  $\leftarrow$  text_info + 1; cur_text < text_ptr; cur_text++)
    if (cur_text-text_link  $\equiv$  macro) {  $\triangleright$  cur_text is the text for a macro  $\triangleleft$ 
      cur_byte  $\leftarrow$  cur_text-tok_start; cur_end  $\leftarrow$  (cur_text + 1)-tok_start; C_printf("%s", "#define_");
      out_state  $\leftarrow$  normal; protect  $\leftarrow$  true;  $\triangleright$  newlines should be preceded by '\n'  $\triangleleft$ 
      while (cur_byte < cur_end) {
        a  $\leftarrow$  *cur_byte++;
        if (cur_byte  $\equiv$  cur_end  $\wedge$  a  $\equiv$  '\n') break;  $\triangleright$  disregard a final newline  $\triangleleft$ 
        if (out_state  $\equiv$  verbatim  $\wedge$  a  $\neq$  string  $\wedge$  a  $\neq$  constant  $\wedge$  a  $\neq$  '\n') C_putc(a);
           $\triangleright$  a high-bit character can occur in a string  $\triangleleft$ 
        else if (a < °200) out_char(a);  $\triangleright$  one-byte token  $\triangleleft$ 
        else {
          a  $\leftarrow$  (a - °200) * °400 + *cur_byte++;
          if (a < °24000) {  $\triangleright$  °24000  $\equiv$  (°250 - °200) * °400  $\triangleleft$ 
            cur_val  $\leftarrow$  a; out_char(identifier);
          }
          else if (a < °50000) {
            confusion("macro_defs_have_strange_char");
          }
          else {
            cur_val  $\leftarrow$  a - °50000; cur_section  $\leftarrow$  cur_val; out_char(section_number);
          }  $\triangleright$  no other cases  $\triangleleft$ 
        }
      }
    }
  protect  $\leftarrow$  false; flush_buffer();
}
```

```
pop_level(false);
```

```
}
```

**55.** A many-way switch is used to send the output. Note that this function is not called if *out\_state*  $\equiv$  *verbatim*, except perhaps with arguments '**\n**' (protect the newline), *string* (end the string), or *constant* (end the constant).

```
static void out_char(eight_bits cur_char)
{
    char *j, *k;    ▷ pointer into byte_mem ◁

restart:
    switch (cur_char) {
    case '\n':
        if (protect  $\wedge$  out_state  $\neq$  verbatim) C_putc('␣');
        if (protect  $\vee$  out_state  $\equiv$  verbatim) C_putc('\\');
        flush_buffer();
        if (out_state  $\neq$  verbatim) out_state  $\leftarrow$  normal;
        break;
    < Case of an identifier 59 >
    < Case of a section number 60 >
    < Cases like != 56 >
    case '=': case '>': C_putc(cur_char); C_putc('␣'); out_state  $\leftarrow$  normal; break;
    case join: out_state  $\leftarrow$  unbreakable; break;
    case constant:
        if (out_state  $\equiv$  verbatim) {
            out_state  $\leftarrow$  num_or_id; break;
        }
        if (out_state  $\equiv$  num_or_id) C_putc('␣');
        out_state  $\leftarrow$  verbatim; break;
    case string:
        if (out_state  $\equiv$  verbatim) out_state  $\leftarrow$  normal;
        else out_state  $\leftarrow$  verbatim;
        break;
    case '/': C_putc('/'); out_state  $\leftarrow$  post_slash; break;
    case '*':
        if (out_state  $\equiv$  post_slash) C_putc('␣');    ▷ fall through ◁
    default: C_putc(cur_char); out_state  $\leftarrow$  normal; break;
    }
}
```

56.  $\langle$  Cases like != 56  $\rangle \equiv$   
**case** *plus\_plus*: *C\_putc*('+'); *C\_putc*('+'); *out\_state*  $\leftarrow$  *normal*; **break**;  
**case** *minus\_minus*: *C\_putc*('−'); *C\_putc*('−'); *out\_state*  $\leftarrow$  *normal*; **break**;  
**case** *minus\_gt*: *C\_putc*('−'); *C\_putc*('>'); *out\_state*  $\leftarrow$  *normal*; **break**;  
**case** *gt\_gt*: *C\_putc*('>'); *C\_putc*('>'); *out\_state*  $\leftarrow$  *normal*; **break**;  
**case** *eq\_eq*: *C\_putc*('='); *C\_putc*('='); *out\_state*  $\leftarrow$  *normal*; **break**;  
**case** *lt\_lt*: *C\_putc*('<'); *C\_putc*('<'); *out\_state*  $\leftarrow$  *normal*; **break**;  
**case** *gt\_eq*: *C\_putc*('>'); *C\_putc*('='); *out\_state*  $\leftarrow$  *normal*; **break**;  
**case** *lt\_eq*: *C\_putc*('<'); *C\_putc*('='); *out\_state*  $\leftarrow$  *normal*; **break**;  
**case** *non\_eq*: *C\_putc*('!'); *C\_putc*('='); *out\_state*  $\leftarrow$  *normal*; **break**;  
**case** *and\_and*: *C\_putc*('&'); *C\_putc*('&'); *out\_state*  $\leftarrow$  *normal*; **break**;  
**case** *or\_or*: *C\_putc*('|'); *C\_putc*('|'); *out\_state*  $\leftarrow$  *normal*; **break**;  
**case** *dot\_dot\_dot*: *C\_putc*('.'); *C\_putc*('.'); *C\_putc*('.'); *out\_state*  $\leftarrow$  *normal*; **break**;  
**case** *colon\_colon*: *C\_putc*(':'); *C\_putc*(':'); *out\_state*  $\leftarrow$  *normal*; **break**;  
**case** *period\_ast*: *C\_putc*('.'); *C\_putc*('\*'); *out\_state*  $\leftarrow$  *normal*; **break**;  
**case** *minus\_gt\_ast*: *C\_putc*('−'); *C\_putc*('>'); *C\_putc*('\*'); *out\_state*  $\leftarrow$  *normal*; **break**;

This code is used in section 55.

57. When an identifier is output to the C file, characters in the range 128–255 must be changed into something else, so the C compiler won't complain. By default, CTANGLE converts the character with code  $16x + y$  to the three characters 'Xxy', but a different transliteration table can be specified. Thus a German might want *grün* to appear as a still readable *gruen*. This makes debugging a lot less confusing.

```
#define translit.length 10
```

```
 $\langle$  Private variables 20  $\rangle + \equiv$   

static char translit[128][translit_length];
```

58.  $\langle$  Set initial values 21  $\rangle + \equiv$   

```
{
    int i;
    for (i  $\leftarrow$  0; i < 128; i++) sprintf(translit[i], "X%02X", (unsigned int)(128 + i));
}
```

59.  $\langle$  Case of an identifier 59  $\rangle \equiv$   
**case** *identifier*:  
**if** (*out\_state*  $\equiv$  *num\_or\_id*) *C\_putc*('␣');  
*j*  $\leftarrow$  (*cur\_val* + *name\_dir*)−*byte\_start*; *k*  $\leftarrow$  (*cur\_val* + *name\_dir* + 1)−*byte\_start*;  
**while** (*j* < *k*) {  
**if** ((*eight\_bits*)(\**j*) < °200) *C\_putc*(\**j*);  
**else** *C\_printf*("s", translit[(*eight\_bits*)(\**j*) − °200]);  
*j*++;  
**}**  
*out\_state*  $\leftarrow$  *num\_or\_id*; **break**;

This code is used in section 55.

60. ⟨ Case of a section number 60 ⟩ ≡

**case** *section\_number*:

```

if (cur_val > 0) C_printf("/*%d:*/", cur_val);
else if (cur_val < 0) C_printf("/*:%d*/", −cur_val);
else if (protect) {
    cur_byte += 4;      ▷ skip line number and file name ◁
    cur_char ← '↵'; goto restart;
}
else {
    sixteen_bits a;
    a ← °400 * *cur_byte++; a += *cur_byte++;      ▷ gets the line number ◁
    C_printf("↵#line%d↵", a); cur_val ← *cur_byte++;
    cur_val ← °400 * (cur_val − °200) + *cur_byte++;      ▷ points to the file name ◁
    for (j ← (cur_val + name_dir)→byte_start, k ← (cur_val + name_dir + 1)→byte_start; j < k; j++) {
        if (*j ≡ '↵' ∨ *j ≡ '') C_putc('↵');
        C_putc(*j);
    }
    C_printf("%s", "↵↵↵");
}
break;
```

This code is used in section 55.

**61. Introduction to the input phase.** We have now seen that CTANGLE will be able to output the full C program, if we can only get that program into the byte memory in the proper format. The input process is something like the output process in reverse, since we compress the text as we read it in and we expand it as we write it out.

There are three main input routines. The most interesting is the one that gets the next token of a C text; the other two are used to scan rapidly past T<sub>E</sub>X text in the CWEB source code. One of the latter routines will jump to the next token that starts with '@', and the other skips to the end of a C comment.

**62.** Control codes in CWEB begin with '@', and the next character identifies the code. Some of these are of interest only to CWEAVE, so CTANGLE ignores them; the others are converted by CTANGLE into internal code numbers by the *ccode* table below. The ordering of these internal code numbers has been chosen to simplify the program logic; larger numbers are given to the control codes that denote more significant milestones.

```
#define ignore 0      ▷ control code of no interest to CTANGLE ◁
#define ord °302      ▷ control code for '@' ◁
#define control_text °303      ▷ control code for '@t', '@^', etc. ◁
#define translit_code °304      ▷ control code for '@l' ◁
#define output_defs_code °305      ▷ control code for '@h' ◁
#define format_code °306      ▷ control code for '@f' ◁
#define definition °307      ▷ control code for '@d' ◁
#define begin_C °310      ▷ control code for '@c' ◁
#define section_name °311      ▷ control code for '@<' ◁
#define new_section °312      ▷ control code for '@_l' and '@*' ◁

⟨Private variables 20⟩ +=
    static eight_bits ccode[256];    ▷ meaning of a char following @ ◁
```

**63.** ⟨Set initial values 21⟩ +=

```
{
    int c;    ▷ must be int so the for loop will end ◁
    for (c ← 0; c < 256; c++) ccode[c] ← ignore;
    ccode['_l'] ← ccode['\t'] ← ccode['\n'] ← ccode['\v'] ← ccode['\r'] ← ccode['\f'] ← ccode['*'] ←
        new_section; ccode['@'] ← '@'; ccode['='] ← string; ccode['d'] ← ccode['D'] ← definition;
    ccode['f'] ← ccode['F'] ← ccode['s'] ← ccode['S'] ← format_code;
    ccode['c'] ← ccode['C'] ← ccode['p'] ← ccode['P'] ← begin_C; ccode['^'] ← ccode[':'] ←
        ccode['.'] ← ccode['t'] ← ccode['T'] ← ccode['q'] ← ccode['Q'] ← control_text;
    ccode['h'] ← ccode['H'] ← output_defs_code; ccode['l'] ← ccode['L'] ← translit_code;
    ccode['&'] ← join; ccode['<'] ← ccode['('] ← section_name; ccode['\''] ← ord;
}
```

64. The *skip\_ahead* procedure reads through the input at fairly high speed until finding the next non-ignorable control code, which it returns.

```
static eight_bits skip_ahead(void)    ▷ skip to next control code ◁
{
  eight_bits c;    ▷ control code found ◁
  while (true) {
    if (loc > limit ∧ (get_line() ≡ false)) return new_section;
    *(limit + 1) ← '@';
    while (*loc ≠ '@') loc++;
    if (loc ≤ limit) {
      loc++; c ← ccode[(eight_bits) *loc]; loc++;
      if (c ≠ ignore ∨ *(loc - 1) ≡ '>') return c;
    }
  }
}
```

65. ⟨Predeclaration of procedures 8⟩ +≡

```
static eight_bits skip_ahead(void);
static boolean skip_comment(boolean);
```

66. The *skip\_comment* procedure reads through the input at somewhat high speed in order to pass over comments, which CTANGLE does not transmit to the output. If the comment is introduced by */\**, *skip\_comment* proceeds until finding the end-comment token *\*/* or a newline; in the latter case *skip\_comment* will be called again by *get\_next*, since the comment is not finished. This is done so that each newline in the C part of a section is copied to the output; otherwise the *#line* commands inserted into the C file by the output routines become useless. On the other hand, if the comment is introduced by *//* (i.e., if it is a C++ “short comment”), it always is simply delimited by the next newline. The boolean argument *is\_long\_comment* distinguishes between the two types of comments.

If *skip\_comment* comes to the end of the section, it prints an error message. No comment, long or short, is allowed to contain '@\_ ' or '@\* '.

⟨Private variables 20⟩ +≡

```
static boolean comment_continues ← false;    ▷ are we scanning a comment? ◁
```

```

67. static boolean skip_comment(    ▷ skips over comments ◁
    boolean is_long_comment)
{
    char c;    ▷ current character ◁
    while (true) {
        if (loc > limit) {
            if (is_long_comment) {
                if (get_line()) return comment_continues ← true;
            }
            else {
                err_print("!_Input_ended_in_mid-comment"); return comment_continues ← false;
            }
        }
        else return comment_continues ← false;
    }
    c ← *(loc++);
    if (is_long_comment ∧ c ≡ '*' ∧ *loc ≡ '/') {
        loc++; return comment_continues ← false;
    }
    if (c ≡ '@') {
        if (ccode[(eight_bits) *loc] ≡ new_section) {
            err_print("!_Section_name_ended_in_mid-comment"); loc--;
            return comment_continues ← false;
        }
        else loc++;
    }
}
}
}

```

**68. Inputting the next token.**

```
#define constant °3
⟨Private variables 20⟩ +=
  static name_pointer cur_section_name;    ▷ name of section just scanned ◁
  static boolean no_where;    ▷ suppress print_where? ◁
```

**69.** As one might expect, *get\_next* consists mostly of a big switch that branches to the various special cases that can arise.

```
#define isxalpha(c) ((c) ≡ '_' ∨ (c) ≡ '$')    ▷ non-alpha characters allowed in identifier ◁
#define ishigh(c) ((eight_bits)(c) > °177)

static eight_bits get_next(void)    ▷ produces the next input token ◁
{
  static boolean preprocessing ← false;
  eight_bits c;    ▷ the current character ◁
  while (true) {
    if (loc > limit) {
      if (preprocessing ∧ *(limit - 1) ≠ '\\') preprocessing ← false;
      if (get_line() ≡ false) return new_section;
      else if (print_where ∧ ¬no_where) {
        print_where ← false; ⟨Insert the line number into tok_mem 85⟩
      }
      else return '\\n';
    }
    c ← *loc;
    if (comment_continues ∨ (c ≡ '/' ∧ (*(loc + 1) ≡ '*' ∨ *(loc + 1) ≡ '/'))) {
      skip_comment(comment_continues ∨ *(loc + 1) ≡ '*');    ▷ scan to end of comment or newline ◁
      if (comment_continues) return '\\n';
      else continue;
    }
    loc++;
    if (xisdigit(c) ∨ c ≡ '.') ⟨Get a constant 73⟩
    else if (c ≡ '\\' ∨ c ≡ '"' ∨ ((c ≡ 'L' ∨ c ≡ 'u' ∨ c ≡ 'U') ∧ (*(loc ≡ '\\' ∨ *loc ≡ '"'))
      ∨ ((c ≡ 'u' ∧ *loc ≡ '8') ∧ (*(loc + 1) ≡ '\\' ∨ *(loc + 1) ≡ '"')))) ⟨Get a string 74⟩
    else if (isalpha(c) ∨ isxalpha(c) ∨ ishigh(c)) ⟨Get an identifier 72⟩
    else if (c ≡ '@') ⟨Get control code and possible section name 75⟩
    else if (xispace(c)) {
      if (¬preprocessing ∨ loc > limit) continue;    ▷ we don't want a blank after a final backslash ◁
      else return '␣';    ▷ ignore spaces and tabs, unless preprocessing ◁
    }
    else if (c ≡ '#' ∧ loc ≡ buffer + 1) preprocessing ← true;
    mistake: ⟨Compress two-symbol operator 71⟩
    return c;
  }
}
```

**70.** ⟨Predeclaration of procedures 8⟩ += static eight\_bits get\_next(void);



**71.** The following code assigns values to the combinations ++, --, ->, >=, <=, ==, <<, >>, !=, and &&, and to the C++ combinations ..., ::, .\* and ->\*. The compound assignment operators (e.g., +=) are treated as separate tokens.

```
#define compress(c) if (loc++ ≤ limit) return c
```

⟨Compress two-symbol operator 71⟩ ≡

```
switch (c) {
case '+':
    if (*loc ≡ '+') compress(plus_plus);
    break;
case '-':
    if (*loc ≡ '-') {
        compress(minus_minus);
    }
    else {
        if (*loc ≡ '>') {
            if *(loc + 1) ≡ '*' {
                loc++; compress(minus_gt_ast);
            }
            else compress(minus_gt);
        }
    }
    break;
case '.':
    if (*loc ≡ '*') {
        compress(period_ast);
    }
    else if (*loc ≡ '.' ∧ *(loc + 1) ≡ '.') {
        loc++; compress(dot_dot_dot);
    }
    break;
case ':':
    if (*loc ≡ ':') compress(colon_colon);
    break;
case '=':
    if (*loc ≡ '=') compress(eq_eq);
    break;
case '>':
    if (*loc ≡ '=') {
        compress(gt_eq);
    }
    else if (*loc ≡ '>') compress(gt_gt);
    break;
case '<':
    if (*loc ≡ '=') {
        compress(lt_eq);
    }
    else if (*loc ≡ '<') compress(lt_lt);
    break;
case '&':
    if (*loc ≡ '&') compress(and_and);
    break;
case '|':
```

```

    if (*loc ≡ ' | ') compress(or_or);
    break;
case '! ':
    if (*loc ≡ '=') compress(non_eq);
    break;
}

```

This code is used in section 69.

72. 〈Get an identifier 72〉 ≡

```

{
    id_first ← --loc;
    while (isalpha((eight_bits) *++loc) ∨ isdigit((eight_bits) *loc)
           ∨ isxalpha((eight_bits) *loc) ∨ ishigh((eight_bits) *loc)) ;
    id_loc ← loc; return identifier;
}

```

This code is used in section 69.

73. 〈Get a constant 73〉 ≡

```

{
    id_first ← loc - 1;
    if (*id_first ≡ '.' ∧ ¬xisdigit(*loc)) goto mistake;    ▷ not a constant ◁
    if (*id_first ≡ '0') {
        if (*loc ≡ 'x' ∨ *loc ≡ 'X') {    ▷ hex constant ◁
            loc++;
            while (xisxdigit(*loc)) loc++;
            goto found;
        }
    }
    while (xisdigit(*loc)) loc++;
    if (*loc ≡ '.' ) {
        loc++;
        while (xisdigit(*loc)) loc++;
    }
    if (*loc ≡ 'e' ∨ *loc ≡ 'E') {    ▷ float constant ◁
        if (*++loc ≡ '+' ∨ *loc ≡ '-') loc++;
        while (xisdigit(*loc)) loc++;
    }
}
found:
while (*loc ≡ 'u' ∨ *loc ≡ 'U' ∨ *loc ≡ 'l' ∨ *loc ≡ 'L' ∨ *loc ≡ 'f' ∨ *loc ≡ 'F') loc++;
id_loc ← loc; return constant;
}

```

This code is used in section 69.

**74.** C strings and character constants, delimited by double and single quotes, respectively, can contain newlines or instances of their own delimiters if they are protected by a backslash. We follow this convention, but do not allow the string to be longer than *longest\_name*.

⟨Get a string 74⟩ ≡

```
{
  char delim ← c;      ▷ what started the string ◁
  id_first ← section_text + 1; id_loc ← section_text; *++id_loc ← delim;
  if (delim ≡ 'L' ∨ delim ≡ 'u' ∨ delim ≡ 'U') {      ▷ wide character constant ◁
    if (delim ≡ 'u' ∧ *loc ≡ '8') {
      *++id_loc ← *loc++;
    }
    delim ← *loc++; *++id_loc ← delim;
  }
  while (true) {
    if (loc ≥ limit) {
      if (*(limit - 1) ≠ '\\') {
        err_print("!_String_didn't_end"); loc ← limit; break;
      }
      if (get_line() ≡ false) {
        err_print("!_Input_ended_in_middle_of_string"); loc ← buffer; break;
      }
      else if (++id_loc ≤ section_text_end) *id_loc ← '\n';      ▷ will print as "\\n" ◁
    }
    if ((c ← *loc++) ≡ delim) {
      if (++id_loc ≤ section_text_end) *id_loc ← c;
      break;
    }
    if (c ≡ '\\') {
      if (loc ≥ limit) continue;
      if (++id_loc ≤ section_text_end) *id_loc ← '\\';
      c ← *loc++;
    }
    if (++id_loc ≤ section_text_end) *id_loc ← c;
  }
  if (id_loc ≥ section_text_end) {
    fputs("\n!_String_too_long:", stdout); term_write(section_text + 1, 25); err_print("...");
  }
  id_loc++; return string;
}
```

This code is used in section 69.

**75.** After an @ sign has been scanned, the next character tells us whether there is more work to do.

⟨Get control code and possible section name 75⟩ ≡

```
{
  c ← ccode[(eight_bits) * loc++];
  switch (c) {
  case ignore: continue;
  case translit_code: err_print("!_Use_@l_in_limbo_only"); continue;
  case control_text:
    while ((c ← skip_ahead()) ≡ '@') ;    ▷ only @@ and @> are expected ◁
    if (*(loc - 1) ≠ '>') err_print("!_Double_@_should_be_used_in_control_text");
    continue;
  case section_name: cur_section_name_char ← *(loc - 1);
    ⟨Scan the section name and make cur_section_name point to it 77⟩
  case string: ⟨Scan a verbatim string 81⟩
  case ord: ⟨Scan an ASCII constant 76⟩
  default: return c;
  }
}
```

This code is cited in section 92.

This code is used in section 69.

**76.** After scanning a valid ASCII constant that follows @', this code plows ahead until it finds the next single quote. (Special care is taken if the quote is part of the constant.) Anything after a valid ASCII constant is ignored; thus, @'\nopq' gives the same result as @'\n'.

⟨Scan an ASCII constant 76⟩ ≡

```
id_first ← loc;
if (*loc ≡ '\\') {
  if (*++loc ≡ '\\') loc++;
}
while (*loc ≠ '\\') {
  if (*loc ≡ '@') {
    if *(loc + 1) ≠ '@' err_print("!_Double_@_should_be_used_in_ASCII_constant");
    else loc++;
  }
  loc++;
  if (loc > limit) {
    err_print("!_String_didn't_end"); loc ← limit - 1; break;
  }
}
loc++; return ord;
```

This code is used in section 75.

```

77.  ⟨Scan the section name and make cur_section_name point to it 77⟩ ≡
{
  char *k;    ▷ pointer into section_text ◁
  ⟨Put section name into section_text 79⟩
  if (k - section_text > 3 ∧ strcmp(k - 2, "...", 3) ≡ 0)
    cur_section_name ← section_lookup(section_text + 1, k - 3, true);    ▷ true means it's a prefix ◁
  else cur_section_name ← section_lookup(section_text + 1, k, false);    ▷ false means it's not ◁
  if (cur_section_name_char ≡ '(')
    ⟨If it's not there, add cur_section_name to the output file stack, or complain we're out of room 47⟩
  return section_name;
}

```

This code is used in section 75.

78. Section names are placed into the *section\_text* array with consecutive spaces, tabs, and carriage-returns replaced by single spaces. There will be no spaces at the beginning or the end. (We set *section\_text*[0] ← '␣' to facilitate this, since the *section\_lookup* routine uses *section\_text*[1] as the first character of the name.)

```

⟨Set initial values 21⟩ +=
  section_text[0] ← '␣';

```

```

79.  ⟨Put section name into section_text 79⟩ ≡
k ← section_text;
while (true) {
  if (loc > limit ∧ get_line() ≡ false) {
    err_print("!␣Input␣ended␣in␣section␣name"); loc ← buffer + 1; break;
  }
  c ← *loc; ⟨If end of name or erroneous nesting, break 80⟩
  loc++;
  if (k < section_text_end) k++;
  if (xisspace(c)) {
    c ← '␣';
    if (*(k - 1) ≡ '␣') k--;
  }
  *k ← c;
}
if (k ≥ section_text_end) {
  fputs("\n!␣Section␣name␣too␣long:␣", stdout); term_write(section_text + 1, 25); printf("...");
  mark_harmless;
}
if (*k ≡ '␣' ∧ k > section_text) k--;

```

This code is used in section 77.

**80.**  $\langle$  If end of name or erroneous nesting, **break** 80  $\rangle \equiv$

```

if ( $c \equiv \text{'@'}$ ) {
   $c \leftarrow *(loc + 1)$ ;
  if ( $c \equiv \text{'>'}$ ) {
     $loc += 2$ ; break;
  }
  if ( $ccode[(\text{eight\_bits})\ c] \equiv new\_section$ ) {
     $err\_print("!\_Section\_name\_didn't\_end")$ ; break;
  }
  if ( $ccode[(\text{eight\_bits})\ c] \equiv section\_name$ ) {
     $err\_print("!\_Nesting\_of\_section\_names\_not\_allowed")$ ; break;
  }
   $*(++k) \leftarrow \text{'@'}$ ;  $loc++$ ;     $\triangleright$  now  $c \equiv *loc$  again  $\triangleleft$ 
}

```

This code is used in section 79.

**81.** At the present point in the program we have  $*(loc - 1) \equiv string$ ; we set  $id\_first$  to the beginning of the string itself, and  $id\_loc$  to its ending-plus-one location in the buffer. We also set  $loc$  to the position just after the ending delimiter.

$\langle$  Scan a verbatim string 81  $\rangle \equiv$

```

{
   $id\_first \leftarrow loc++$ ;  $*(limit + 1) \leftarrow \text{'@'}$ ;  $*(limit + 2) \leftarrow \text{'>'}$ ;
  while ( $*loc \neq \text{'@'} \vee *(loc + 1) \neq \text{'>'}$ )  $loc++$ ;
  if ( $loc \geq limit$ )  $err\_print("!\_Verbatim\_string\_didn't\_end")$ ;
   $id\_loc \leftarrow loc$ ;  $loc += 2$ ; return  $string$ ;
}

```

This code is used in section 75.

**82. Scanning a macro definition.** The rules for generating the replacement texts corresponding to macros and C texts of a section are almost identical; the only differences are that

- Section names are not allowed in macros; in fact, the appearance of a section name terminates such macros and denotes the name of the current section.
- The symbols `@d` and `@f` and `@c` are not allowed after section names, while they terminate macro definitions.
- Spaces are inserted after right parentheses in macros, because the ANSI C preprocessor sometimes requires it.

Therefore there is a single procedure *scan\_repl* whose parameter *t* specifies either *macro* or *section\_name*. After *scan\_repl* has acted, *cur\_text* will point to the replacement text just generated, and *next\_control* will contain the control code that terminated the activity.

```
#define macro 0
#define app_repl(c)
{
  if (tok_ptr == tok_mem_end) overflow("token");
  *tok_ptr++ <- c;
}
```

⟨Private variables 20⟩ +≡

```
static text_pointer cur_text;    ▷ replacement text formed by scan_repl ◁
static eight_bits next_control;
```

**83. static void scan\_repl(** ▷ creates a replacement text ◁  
**eight\_bits t)**

```
{
  sixteen_bits a;    ▷ the current token ◁
  if (t == section_name) {
    ⟨Insert the line number into tok_mem 85⟩
  }
  while (true)
    switch (a <- get_next()) {
      ⟨In cases that a is a non-char token (identifier, section_name, etc.), either process it and change
        a to a byte that should be stored, or continue if a should be ignored, or goto done if a
        signals the end of this replacement text 86⟩
      case ' ': app_repl(a);
        if (t == macro) app_repl('␣');
        break;
      default: app_repl(a);    ▷ store a in tok_mem ◁
    }
  done: next_control <- (eight_bits) a;
  if (text_ptr > text_info_end) overflow("text");
  cur_text <- text_ptr; (++text_ptr)-tok_start <- tok_ptr;
}
```

**84. ⟨Predeclaration of procedures 8⟩ +≡ static void scan\_repl(eight\_bits);**

**85.** Here is the code for the line number: first a **sixteen\_bits** equal to  $^{\circ}150000$ ; then the numeric line number; then a pointer to the file name.

```

⟨Insert the line number into tok_mem 85⟩ ≡
  store_two_bytes( $^{\circ}150000$ );
  if (changing ∧ include_depth ≡ change_depth) {    ▷ correction made Feb 2017 ◁
    id_first ← change_file_name; store_two_bytes((sixteen_bits) change_line);
  } else {
    id_first ← cur_file_name; store_two_bytes((sixteen_bits) cur_line);
  }
  id_loc ← id_first + strlen(id_first);
  {
    int a_l ← id_lookup(id_first, id_loc, 0) - name_dir;
    app_repl((a_l /  $^{\circ}400$ ) +  $^{\circ}200$ ); app_repl(a_l %  $^{\circ}400$ );
  }

```

This code is used in sections 69, 83, and 86.

**86.** ⟨In cases that *a* is a non-**char** token (*identifier*, *section\_name*, etc.), either process it and change *a* to a byte that should be stored, or **continue** if *a* should be ignored, or **goto done** if *a* signals the end of this replacement text 86⟩ ≡

```

case identifier: a ← id_lookup(id_first, id_loc, 0) - name_dir; app_repl((a /  $^{\circ}400$ ) +  $^{\circ}200$ );
  app_repl(a %  $^{\circ}400$ ); break;
case section_name:
  if (t ≠ section_name) goto done;
  else {
    ⟨Was an '@' missed here? 87⟩
    a ← cur_section_name - name_dir; app_repl((a /  $^{\circ}400$ ) +  $^{\circ}250$ ); app_repl(a %  $^{\circ}400$ );
    ⟨Insert the line number into tok_mem 85⟩
    break;
  }
case output_defs_code:
  if (t ≠ section_name) err_print("!_Misplaced_@h");
  else {
    output_defs_seen ← true; a ← output_defs_flag; app_repl((a /  $^{\circ}400$ ) +  $^{\circ}200$ ); app_repl(a %  $^{\circ}400$ );
    ⟨Insert the line number into tok_mem 85⟩
  }
  break;
case constant: case string: ⟨Copy a string or verbatim construction or numerical constant 88⟩
case ord: ⟨Copy an ASCII constant 89⟩
case definition: case format_code: case begin_C:
  if (t ≠ section_name) goto done;
  else {
    err_print("!_@d, _@f _and _@c _are _ignored _in _C _text"); continue;
  }
case new_section: goto done;

```

This code is used in section 83.



87.  $\langle$  Was an '@' missed here? 87  $\rangle \equiv$

```
{
  char *try_loc ← loc;
  while (*try_loc ≡ '␣' ∧ try_loc < limit) try_loc++;
  if (*try_loc ≡ '+' ∧ try_loc < limit) try_loc++;
  while (*try_loc ≡ '␣' ∧ try_loc < limit) try_loc++;
  if (*try_loc ≡ '=') err_print("!␣Missing␣'@␣'␣before␣a␣named␣section");
    ▷ user who isn't defining a section should put newline after the name, as explained in the manual ◁
}
```

This code is used in section 86.

88.  $\langle$  Copy a string or verbatim construction or numerical constant 88  $\rangle \equiv$

```
app_repl(a);    ▷ string or constant ◁
while (id_first < id_loc) {    ▷ simplify @@ pairs ◁
  if (*id_first ≡ '@') {
    if (*(id_first + 1) ≡ '@') id_first++;
    else err_print("!␣Double␣@␣should␣be␣used␣in␣string");
  }
  app_repl(*id_first++);
}
app_repl(a); break;
```

This code is used in section 86.

89. This section should be rewritten on machines that don't use ASCII code internally.

⟨ Copy an ASCII constant 89 ⟩ ≡

```
{
  int c ← (eight_bits) *id_first;
  if (c ≡ '\\') {
    c ← *++id_first;
    if (c ≥ '0' ∧ c ≤ '7') {
      c -= '0';
      if (*(id_first + 1) ≥ '0' ∧ *(id_first + 1) ≤ '7') {
        c ← 8 * c + *(++id_first) - '0';
        if (*(id_first + 1) ≥ '0' ∧ *(id_first + 1) ≤ '7' ∧ c < 32) c ← 8 * c + *(++id_first) - '0';
      }
    }
  }
  else
    switch (c) {
      case 't': c ← '\t'; break;
      case 'n': c ← '\n'; break;
      case 'b': c ← '\b'; break;
      case 'f': c ← '\f'; break;
      case 'v': c ← '\v'; break;
      case 'r': c ← '\r'; break;
      case 'a': c ← '\7'; break;
      case '?': c ← '?'; break;
      case 'x':
        if (xisdigit(*(id_first + 1))) c ← *(++id_first) - '0';
        else if (xisdigit(*(id_first + 1))) {
          ++id_first; c ← toupper((eight_bits) *id_first) - 'A' + 10;
        }
        if (xisdigit(*(id_first + 1))) c ← 16 * c + *(++id_first) - '0';
        else if (xisdigit(*(id_first + 1))) {
          ++id_first; c ← 16 * c + toupper((eight_bits) *id_first) - 'A' + 10;
        }
        break;
      case '\\': c ← '\\'; break;
      case '\': c ← '\'; break;
      case '\"': c ← '\"'; break;
      default: err_print("!_Unrecognized_escape_sequence");
    }
} ▷ at this point c should have been converted to its ASCII code number ◁
app_repl(constant);
if (c ≥ 100) app_repl('0' + c/100);
if (c ≥ 10) app_repl('0' + (c/10) % 10);
app_repl('0' + c % 10); app_repl(constant);
}
break;
```

This code is used in section 86.

**90. Scanning a section.** The *scan\_section* procedure starts when ‘@\_’ or ‘@\*’ has been sensed in the input, and it proceeds until the end of that section. It uses *section\_count* to keep track of the current section number; with luck, CWEAVE and CTANGLE will both assign the same numbers to sections.

The body of *scan\_section* is a loop where we look for control codes that are significant to CTANGLE: those that delimit a definition, the C part of a module, or a new module.

```
static void scan_section(void)
{
    name_pointer p;    ▷ section name for the current section ◁
    text_pointer q;    ▷ text for the current section ◁
    sixteen_bits a;    ▷ token for left-hand side of definition ◁
    section_count++; no_where ← true;
    if (*(loc - 1) ≡ '*' ∧ show_progress) {    ▷ starred section ◁
        printf("%d", section_count); update_terminal;
    }
    next_control ← 0;
    while (true) {
        ◁ Skip ahead until next_control corresponds to @d, @<, @_ or the like 92 ▷
        if (next_control ≡ definition) {    ▷ @d ◁
            ◁ Scan a definition 93 ▷
            continue;
        }
        if (next_control ≡ begin_C) {    ▷ @c or @p ◁
            p ← name_dir; break;
        }
        if (next_control ≡ section_name) {    ▷ @< or @( ◁
            p ← cur_section_name; ◁ If section is not being defined, continue 94 ▷
            break;
        }
        return;    ▷ @_ or @* ◁
    }
    no_where ← print_where ← false; ◁ Scan the C part of the current section 95 ▷
}
```

**91.** ◁ Predeclaration of procedures 8 ▷ +≡ **static void scan\_section(void);**

**92.** At the top of this loop, if *next\_control* ≡ *section\_name*, the section name has already been scanned (see ◁ Get control code and possible section name 75 ▷). Thus, if we encounter *next\_control* ≡ *section\_name* in the skip-ahead process, we should likewise scan the section name, so later processing will be the same in both cases.

```
◁ Skip ahead until next_control corresponds to @d, @<, @_ or the like 92 ▷ ≡
while (next_control < definition)    ▷ definition is the lowest of the "significant" codes ◁
    if ((next_control ← skip_ahead()) ≡ section_name) {
        loc -= 2; next_control ← get_next();
    }
```

This code is used in section 90.

93.  $\langle$  Scan a definition 93  $\rangle \equiv$

```
{
  while ((next_control ← get_next()) ≡ '\n') ;    ▷ allow newline before definition ◁
  if (next_control ≠ identifier) {
    err_print("!_Definition_flushed,_must_start_with_identifier"); continue;
  }
  app_repl((((a ← id_lookup(id_first, id_loc, 0) - name_dir) / °400) + °200);    ▷ append the lhs ◁
  app_repl(a % °400);
  if (*loc ≠ '(') {    ▷ identifier must be separated from replacement text ◁
    app_repl(string); app_repl(' '); app_repl(string);
  }
  scan_repl(macro); cur_text→text_link ← macro;
}
```

This code is used in section 90.

94. If the section name is not followed by = or +=, no C code is forthcoming: the section is being cited, not being defined. This use is illegal after the definition part of the current section has started, except inside a comment, but CTANGLE does not enforce this rule; it simply ignores the offending section name and everything following it, up to the next significant control code.

$\langle$  If section is not being defined, **continue** 94  $\rangle \equiv$

```
while ((next_control ← get_next()) ≡ '+') ;    ▷ allow optional += ◁
if (next_control ≠ '=' ∧ next_control ≠ eq_eq) continue;
```

This code is used in section 90.

95.  $\langle$  Scan the C part of the current section 95  $\rangle \equiv$

$\langle$  Insert the section number into *tok\_mem* 96  $\rangle$

*scan\_repl*(*section\_name*); ▷ now *cur\_text* points to the replacement text ◁

$\langle$  Update the data structure so that the replacement text is accessible 97  $\rangle$

This code is used in section 90.

96.  $\langle$  Insert the section number into *tok\_mem* 96  $\rangle \equiv$

*store\_two\_bytes*((**sixteen\_bits**)(°150000 + *section\_count*)); ▷ °150000 ≡ °320 \* °400 ◁

This code is used in section 95.

97.  $\langle$  Update the data structure so that the replacement text is accessible 97  $\rangle \equiv$

```
if (p ≡ name_dir ∨ p ≡ Λ) {    ▷ unnamed section, or bad section name ◁
  (last_unnamed)→text_link ← cur_text - text_info; last_unnamed ← cur_text;
}
else if (p→equiv ≡ (void *) text_info) p→equiv ← (void *) cur_text;    ▷ first section of this name ◁
else {
  q ← (text_pointer) p→equiv;
  while (q→text_link < section_flag) q ← q→text_link + text_info;    ▷ find end of list ◁
  q→text_link ← cur_text - text_info;
}
cur_text→text_link ← section_flag;    ▷ mark this replacement text as a nonmacro ◁
```

This code is used in section 95.

98. **static void** *phase\_one*(**void**)

```
{
  phase ← 1; section_count ← 0; reset_input(); skip_limbo();
  while (¬input_has_ended) scan_section();
  check_complete(); phase ← 2;
}
```

99. ⟨Predeclaration of procedures 8⟩ +≡ **static void** *phase\_one*(**void**);

100. Only a small subset of the control codes is legal in limbo, so limbo processing is straightforward.

**static void** *skip\_limbo*(**void**)

```
{
  char c;
  while (true) {
    if (loc > limit ∧ get_line() ≡ false) return;
    *(limit + 1) ← '@';
    while (*loc ≠ '@') loc++;
    if (loc++ ≤ limit) {
      c ← *loc++;
      if (ccode[(eight_bits) c] ≡ new_section) break;
      switch (ccode[(eight_bits) c]) {
        case translit_code: ⟨Read in transliteration of a character 102⟩
          break;
        case format_code: case '@': break;
        case control_text:
          if (c ≡ 'q' ∨ c ≡ 'Q') {
            while ((c ← skip_ahead()) ≡ '@') ;
            if (*(loc - 1) ≠ '>') err_print("!Double_@_should_be_used_in_control_text");
            break;
          } ▷ otherwise fall through ◁
        default: err_print("!Double_@_should_be_used_in_limbo");
      }
    }
  }
}
```

101. ⟨Predeclaration of procedures 8⟩ +≡ **static void** *skip\_limbo*(**void**);

```

102.  ⟨ Read in transliteration of a character 102 ⟩ ≡
  while (xisspace(*loc) ∧ loc < limit) loc++;
  loc += 3;
  if (loc > limit ∨ ¬xisxdigit(*(loc - 3)) ∨ ¬xisxdigit(*(loc - 2))
      ∨ (*(loc - 3) ≥ '0' ∧ *(loc - 3) ≤ '7') ∨ ¬xisspace(*(loc - 1)))
    err_print("! Improper hex number following @1");
  else {
    unsigned int i;
    char *beg;
    sscanf(loc - 3, "%x", &i);
    while (xisspace(*loc) ∧ loc < limit) loc++;
    beg ← loc;
    while (loc < limit ∧ (xisalpha(*loc) ∨ xisdigit(*loc) ∨ *loc ≡ '_')) loc++;
    if (loc - beg ≥ translit_length) err_print("! Replacement string in @1 too long");
    else {
      strncpy(translit[i - °200], beg, (size_t)(loc - beg)); translit[i - °200][loc - beg] ← '\0';
    }
  }
}

```

This code is used in section 100.

**103.** Because on some systems the difference between two pointers is a `ptrdiff_t` but not an `int`, we use `%ld` to print these quantities.

```

void print_stats(void)
{
  puts("\nMemory usage statistics:");
  printf("%ld names (out of %ld)\n", (ptrdiff_t)(name_ptr - name_dir), (long) max_names);
  printf("%ld replacement texts (out of %ld)\n", (ptrdiff_t)(text_ptr - text_info), (long)
    max_texts);
  printf("%ld bytes (out of %ld)\n", (ptrdiff_t)(byte_ptr - byte_mem), (long) max_bytes);
  printf("%ld tokens (out of %ld)\n", (ptrdiff_t)(tok_ptr - tok_mem), (long) max_toks);
}

```

**104. Index.** Here is a cross-reference table for CTANGLE. All sections in which an identifier is used are listed with that identifier, except that reserved words are indexed only when they appear in format definitions, and the appearances of identifiers in section names are not indexed. Underlined entries correspond to where the identifier was declared. Error messages and a few other things like “ASCII code dependencies” are indexed here too.

@d, @f and @c are ignored in C text: [86](#).

a: [38](#), [54](#), [60](#), [83](#), [90](#).

a\_l: [85](#).

ac: [2](#), [14](#).

active\_file: [15](#).

an\_output\_file: [45](#), [47](#), [50](#).

and\_and: [5](#), [56](#), [71](#).

app\_repl: [82](#), [83](#), [85](#), [86](#), [88](#), [89](#), [93](#).

argc: [2](#), [14](#).

argv: [2](#), [14](#).

ASCII code dependencies: [5](#), [28](#), [89](#).

av: [2](#), [14](#).

banner: [1](#), [2](#).

beg: [102](#).

begin\_C: [62](#), [63](#), [86](#), [90](#).

bool: [4](#).

boolean: [3](#), [7](#), [8](#), [9](#), [11](#), [14](#), [24](#), [35](#), [36](#), [42](#), [52](#),  
[65](#), [66](#), [67](#), [68](#), [69](#).

buf\_size: [17](#).

buffer: [6](#), [69](#), [74](#), [79](#).

buffer\_end: [6](#).

byte\_field: [31](#), [32](#).

byte\_mem: [10](#), [19](#), [55](#), [103](#).

byte\_mem\_end: [10](#).

byte\_ptr: [10](#), [103](#).

byte\_start: [10](#), [24](#), [31](#), [59](#), [60](#).

c: [63](#), [64](#), [67](#), [69](#), [89](#), [100](#).

C\_file: [14](#), [15](#), [50](#).

C\_file\_name: [14](#), [48](#).

C\_printf: [15](#), [54](#), [59](#), [60](#).

C\_putc: [15](#), [38](#), [43](#), [54](#), [55](#), [56](#), [59](#), [60](#).

Cannot open output file: [50](#).

ccode: [62](#), [63](#), [64](#), [67](#), [75](#), [80](#), [100](#).

change\_depth: [7](#), [85](#).

change\_file: [7](#).

change\_file\_name: [7](#), [85](#).

change\_line: [7](#), [85](#).

change\_pending: [9](#).

changed\_section: [9](#).

changing: [7](#), [85](#).

check\_complete: [8](#), [98](#).

colon\_colon: [5](#), [56](#), [71](#).

comment\_continues: [66](#), [67](#), [69](#).

common\_init: [2](#), [16](#).

compress: [71](#).

confusion: [12](#), [54](#).

constant: [38](#), [54](#), [55](#), [68](#), [73](#), [86](#), [88](#), [89](#).

control\_text: [62](#), [63](#), [75](#), [100](#).

ctangle: [2](#), [3](#).

cur\_byte: [31](#), [32](#), [33](#), [34](#), [36](#), [38](#), [50](#), [54](#), [60](#).

cur\_char: [55](#), [60](#).

cur\_end: [31](#), [32](#), [33](#), [34](#), [36](#), [38](#), [50](#), [54](#).

cur\_file: [7](#).

cur\_file\_name: [7](#), [85](#).

cur\_line: [7](#), [43](#), [48](#), [50](#), [85](#).

cur\_name: [31](#), [32](#), [33](#), [34](#), [50](#).

cur\_out\_file: [45](#), [46](#), [47](#), [48](#), [50](#).

cur\_repl: [31](#), [32](#), [33](#), [34](#), [36](#), [50](#).

cur\_section: [31](#), [32](#), [33](#), [34](#), [38](#), [54](#).

cur\_section\_name: [47](#), [68](#), [77](#), [86](#), [90](#).

cur\_section\_name\_char: [45](#), [75](#), [77](#).

cur\_state: [32](#), [34](#), [36](#).

cur\_text: [54](#), [82](#), [83](#), [93](#), [95](#), [97](#).

cur\_val: [37](#), [38](#), [54](#), [59](#), [60](#).

cweave: [3](#).

definition: [62](#), [63](#), [86](#), [90](#), [92](#).

Definition flushed...: [93](#).

delim: [74](#).

done: [83](#), [86](#).

dot\_dot\_dot: [5](#), [56](#), [71](#).

Double @ should be used...: [75](#), [76](#), [88](#), [100](#).

dummy: [10](#).

eight\_bits: [3](#), [6](#), [11](#), [19](#), [20](#), [24](#), [25](#), [31](#), [42](#), [53](#),  
[55](#), [59](#), [62](#), [64](#), [65](#), [67](#), [69](#), [70](#), [72](#), [75](#), [80](#),  
[82](#), [83](#), [84](#), [89](#), [100](#).

end\_field: [31](#), [32](#).

end\_output\_files: [45](#), [46](#), [47](#), [48](#), [50](#).

eq\_eq: [5](#), [56](#), [71](#), [94](#).

equiv: [22](#), [25](#), [34](#), [40](#), [50](#), [97](#).

equiv\_or\_xref: [10](#), [22](#).

err\_print: [13](#), [40](#), [67](#), [74](#), [75](#), [76](#), [79](#), [80](#), [81](#), [86](#),  
[87](#), [88](#), [89](#), [93](#), [100](#), [102](#).

error\_message: [12](#).

exit: [4](#).

false: [3](#), [4](#), [24](#), [36](#), [48](#), [52](#), [54](#), [64](#), [66](#), [67](#), [69](#),  
[74](#), [77](#), [79](#), [90](#), [100](#).

fatal: [12](#), [13](#), [50](#).

fatal\_message: [12](#).

fclose: [50](#).

fflush: [15](#).

file: [7](#).

file\_name: [7](#).

first: [24](#).

flag: [36](#).

*flags*: [14](#).  
*flush\_buffer*: [43](#), [44](#), [48](#), [50](#), [54](#), [55](#).  
*fopen*: [50](#).  
*format\_code*: [62](#), [63](#), [86](#), [100](#).  
*found*: [73](#).  
*fprintf*: [15](#).  
*fputs*: [40](#), [48](#), [74](#), [79](#).  
*fwrite*: [15](#).  
*get\_line*: [8](#), [64](#), [67](#), [69](#), [74](#), [79](#), [100](#).  
*get\_next*: [66](#), [69](#), [70](#), [83](#), [92](#), [93](#), [94](#).  
*get\_output*: [37](#), [38](#), [39](#), [41](#), [48](#), [50](#).  
*getenv*: [4](#).  
*gt\_eq*: [5](#), [56](#), [71](#).  
*gt\_gt*: [5](#), [56](#), [71](#).  
*h*: [10](#).  
*harmless\_message*: [12](#).  
*hash*: [10](#).  
*hash\_end*: [10](#).  
**hash\_pointer**: [10](#).  
 high-bit character handling: [38](#), [54](#), [59](#), [69](#).  
*history*: [12](#), [13](#).  
*i*: [58](#), [102](#).  
*id\_first*: [5](#), [72](#), [73](#), [74](#), [76](#), [81](#), [85](#), [86](#), [88](#), [89](#), [93](#).  
*id\_loc*: [5](#), [72](#), [73](#), [74](#), [81](#), [85](#), [86](#), [88](#), [93](#).  
*id\_lookup*: [11](#), [85](#), [86](#), [93](#).  
*identifier*: [37](#), [38](#), [54](#), [59](#), [72](#), [86](#), [93](#).  
*idx\_file*: [14](#), [15](#).  
*idx\_file\_name*: [14](#).  
*ignore*: [62](#), [63](#), [64](#), [75](#).  
*Ilk*: [10](#).  
 Improper hex number...: [102](#).  
*include\_depth*: [7](#), [85](#).  
*init\_node*: [11](#), [23](#), [25](#).  
*init\_p*: [11](#), [25](#).  
 Input ended in mid-comment: [67](#).  
 Input ended in middle of string: [74](#).  
 Input ended in section name: [79](#).  
*input\_has\_ended*: [7](#), [98](#).  
*is\_long\_comment*: [66](#), [67](#).  
*isalpha*: [4](#), [6](#), [69](#), [72](#).  
*isdigit*: [4](#), [6](#), [72](#).  
*ishigh*: [69](#), [72](#).  
*islower*: [6](#).  
*isspace*: [6](#).  
*isupper*: [6](#).  
*isxalpha*: [69](#), [72](#).  
*isxdigit*: [6](#).  
*j*: [55](#).  
*join*: [28](#), [55](#), [63](#).  
*k*: [55](#), [77](#).  
*l*: [24](#).  
*last\_unnamed*: [26](#), [27](#), [97](#).

*length*: [10](#), [24](#).  
*limit*: [6](#), [64](#), [67](#), [69](#), [71](#), [74](#), [76](#), [79](#), [81](#), [87](#), [100](#), [102](#).  
*line*: [7](#).  
**#line**: [60](#).  
*link*: [10](#).  
*llink*: [10](#).  
*loc*: [6](#), [64](#), [67](#), [69](#), [71](#), [72](#), [73](#), [74](#), [75](#), [76](#), [79](#), [80](#),  
     [81](#), [87](#), [90](#), [92](#), [93](#), [100](#), [102](#).  
*long\_buf\_size*: [17](#).  
*longest\_name*: [17](#), [45](#), [74](#).  
*lt\_eq*: [5](#), [56](#), [71](#).  
*lt\_lt*: [5](#), [56](#), [71](#).  
*macro*: [27](#), [48](#), [54](#), [82](#), [83](#), [93](#).  
*main*: [2](#), [14](#).  
*make\_xrefs*: [14](#).  
*mark\_error*: [12](#).  
*mark\_harmless*: [12](#), [48](#), [79](#).  
*max\_bytes*: [17](#), [103](#).  
*max\_file\_name\_length*: [7](#).  
*max\_files*: [45](#), [46](#).  
*max\_include\_depth*: [7](#).  
*max\_names*: [17](#), [103](#).  
*max\_sections*: [17](#).  
*max\_texts*: [17](#), [20](#), [26](#), [103](#).  
*max\_toks*: [17](#), [20](#), [103](#).  
*minus\_gt*: [5](#), [56](#), [71](#).  
*minus\_gt\_ast*: [5](#), [56](#), [71](#).  
*minus\_minus*: [5](#), [56](#), [71](#).  
 Misplaced @h: [86](#).  
 Missing '@'...: [87](#).  
*mistake*: [69](#), [73](#).  
*name\_dir*: [10](#), [23](#), [33](#), [40](#), [59](#), [60](#), [85](#), [86](#), [90](#),  
     [93](#), [97](#), [103](#).  
*name\_dir\_end*: [10](#).  
*name\_field*: [31](#), [32](#).  
**name\_info**: [10](#).  
**name\_pointer**: [10](#), [11](#), [24](#), [25](#), [31](#), [34](#), [35](#),  
     [45](#), [68](#), [90](#).  
*name\_ptr*: [10](#), [103](#).  
*names\_match*: [11](#), [24](#).  
 Nesting of section names...: [80](#).  
*new\_line*: [15](#), [48](#).  
*new\_section*: [62](#), [63](#), [64](#), [67](#), [69](#), [80](#), [86](#), [100](#).  
*next\_control*: [82](#), [83](#), [90](#), [92](#), [93](#), [94](#).  
 No program text...: [48](#).  
*no\_where*: [68](#), [69](#), [90](#).  
*node*: [25](#).  
*non\_eq*: [5](#), [56](#), [71](#).  
*normal*: [42](#), [54](#), [55](#), [56](#).  
 Not present: <section name>: [40](#).  
*num\_or\_id*: [42](#), [55](#), [59](#).  
*or\_or*: [5](#), [56](#), [71](#).



- ord*: [62](#), [63](#), [75](#), [76](#), [86](#).
- out\_char*: [37](#), [38](#), [53](#), [54](#), [55](#).
- out\_state*: [38](#), [42](#), [54](#), [55](#), [56](#), [59](#).
- output\_defs*: [34](#), [36](#), [38](#), [51](#), [53](#), [54](#).
- output\_defs\_code*: [62](#), [63](#), [86](#).
- output\_defs\_flag*: [28](#), [38](#), [86](#).
- output\_defs\_seen*: [51](#), [52](#), [86](#).
- output\_file\_name*: [45](#), [50](#).
- output\_files*: [45](#), [46](#), [47](#).
- output\_state**: [31](#), [32](#).
- overflow*: [13](#), [29](#), [34](#), [47](#), [82](#), [83](#).
- p*: [24](#), [25](#), [34](#), [90](#).
- period\_ast*: [5](#), [56](#), [71](#).
- phase*: [3](#), [98](#).
- phase\_one*: [2](#), [98](#), [99](#).
- phase\_two*: [2](#), [48](#), [49](#).
- plus\_plus*: [5](#), [56](#), [71](#).
- pop\_level*: [35](#), [36](#), [38](#), [54](#).
- post\_slash*: [42](#), [55](#).
- preprocessing*: [69](#).
- print\_id*: [10](#).
- print\_prefix\_name*: [11](#).
- print\_section\_name*: [11](#), [40](#).
- print\_stats*: [16](#), [103](#).
- print\_where*: [9](#), [68](#), [69](#), [90](#).
- printf*: [4](#), [43](#), [48](#), [50](#), [79](#), [90](#), [103](#).
- program*: [2](#), [3](#).
- protect*: [42](#), [54](#), [55](#), [60](#).
- ptrdiff\_t**: [4](#).
- push\_level*: [34](#), [35](#), [40](#), [54](#).
- putc*: [15](#).
- putchar*: [15](#).
- puts*: [2](#), [103](#).
- putxchar*: [15](#).
- q*: [90](#).
- repl\_field*: [31](#), [32](#).
- Replacement string in @l...: [102](#).
- reset\_input*: [8](#), [98](#).
- restart*: [38](#), [40](#), [55](#), [60](#).
- Rlink*: [10](#).
- rlink*: [10](#).
- root*: [10](#).
- scan\_repl*: [82](#), [83](#), [84](#), [93](#), [95](#).
- scan\_section*: [90](#), [91](#), [98](#).
- scn\_file*: [14](#), [15](#).
- scn\_file\_name*: [14](#).
- Section name didn't end: [80](#).
- Section name ended in mid-comment: [67](#).
- Section name too long: [79](#).
- section\_count*: [9](#), [90](#), [96](#), [98](#).
- section\_field*: [31](#), [32](#).
- section\_flag*: [26](#), [36](#), [97](#).
- section\_lookup*: [11](#), [77](#), [78](#).
- section\_name*: [62](#), [63](#), [75](#), [77](#), [80](#), [82](#), [83](#), [86](#), [90](#), [92](#), [95](#).
- section\_number*: [37](#), [38](#), [54](#), [60](#).
- section\_text*: [5](#), [74](#), [77](#), [78](#), [79](#).
- section\_text\_end*: [5](#), [74](#), [79](#).
- show\_banner*: [2](#), [14](#).
- show\_happiness*: [14](#), [48](#).
- show\_progress*: [14](#), [43](#), [48](#), [50](#), [90](#).
- show\_stats*: [14](#).
- sixteen\_bits**: [3](#), [9](#), [19](#), [29](#), [30](#), [31](#), [38](#), [54](#), [60](#), [83](#), [85](#), [90](#), [96](#).
- skip\_ahead*: [64](#), [65](#), [75](#), [92](#), [100](#).
- skip\_comment*: [65](#), [66](#), [67](#), [69](#).
- skip\_limbo*: [98](#), [100](#), [101](#).
- spotless*: [12](#).
- sprint\_section\_name*: [11](#), [50](#).
- sprintf*: [58](#).
- sscanf*: [102](#).
- stack*: [31](#), [32](#), [33](#), [36](#), [38](#), [48](#), [50](#).
- stack\_end*: [32](#), [34](#).
- stack\_pointer**: [31](#), [32](#).
- stack\_ptr*: [31](#), [32](#), [33](#), [34](#), [36](#), [38](#), [48](#), [50](#).
- stack\_size*: [17](#), [32](#).
- stdout*: [15](#), [40](#), [48](#), [74](#), [79](#).
- store\_two\_bytes*: [29](#), [30](#), [85](#), [96](#).
- strcmp*: [4](#).
- string*: [28](#), [38](#), [54](#), [55](#), [63](#), [74](#), [75](#), [81](#), [86](#), [88](#), [93](#).
- String didn't end: [74](#), [76](#).
- String too long: [74](#).
- strlen*: [4](#), [85](#).
- strncmp*: [24](#), [77](#).
- strncpy*: [102](#).
- system dependencies: [34](#).
- t*: [24](#), [25](#), [83](#).
- term\_write*: [10](#), [15](#), [74](#), [79](#).
- tex\_file*: [14](#), [15](#).
- tex\_file\_name*: [14](#).
- text**: [19](#), [20](#).
- text\_info*: [19](#), [20](#), [21](#), [22](#), [25](#), [26](#), [27](#), [33](#), [36](#), [40](#), [48](#), [54](#), [97](#), [103](#).
- text\_info\_end*: [20](#), [83](#).
- text\_link*: [19](#), [26](#), [27](#), [33](#), [36](#), [48](#), [54](#), [93](#), [97](#).
- text\_pointer**: [19](#), [20](#), [26](#), [31](#), [34](#), [50](#), [82](#), [90](#), [97](#).
- text\_ptr*: [19](#), [20](#), [21](#), [54](#), [83](#), [103](#).
- This can't happen: [12](#).
- tok\_mem*: [2](#), [19](#), [20](#), [21](#), [26](#), [29](#), [31](#), [32](#), [83](#), [103](#).
- tok\_mem\_end*: [20](#), [29](#), [82](#).
- tok\_ptr*: [19](#), [20](#), [21](#), [29](#), [82](#), [83](#), [103](#).
- tok\_start*: [19](#), [21](#), [26](#), [31](#), [33](#), [34](#), [36](#), [50](#), [54](#), [83](#).
- toupper*: [89](#).
- translit*: [57](#), [58](#), [59](#), [102](#).

*translit\_code*: [62](#), [63](#), [75](#), [100](#).  
*translit\_length*: [57](#), [102](#).  
*true*: [3](#), [4](#), [38](#), [42](#), [54](#), [64](#), [67](#), [69](#), [74](#), [77](#), [79](#),  
[83](#), [86](#), [90](#), [100](#).  
*try\_loc*: [87](#).  
**uint16\_t**: [3](#), [4](#).  
**uint8\_t**: [3](#), [4](#).  
*unbreakable*: [42](#), [55](#).  
Unrecognized escape sequence: [89](#).  
*update\_terminal*: [15](#), [43](#), [48](#), [50](#), [90](#).  
Use @l in limbo...: [75](#).  
*verbatim*: [38](#), [42](#), [54](#), [55](#).  
Verbatim string didn't end: [81](#).  
*web\_file*: [7](#).  
*web\_file\_name*: [7](#).  
*web\_file\_open*: [7](#), [48](#).  
*wrap\_up*: [2](#), [13](#).  
*writeloop*: [48](#).  
Writing the output...: [48](#).  
*x*: [29](#).  
*xisalpha*: [6](#), [102](#).  
*xisdigit*: [6](#), [69](#), [73](#), [89](#), [102](#).  
*xislower*: [6](#).  
*xisspace*: [6](#), [69](#), [79](#), [102](#).  
*xisupper*: [6](#).  
*xisxdigit*: [6](#), [73](#), [89](#), [102](#).

- ⟨ Case of a section number 60 ⟩ Used in section 55.
- ⟨ Case of an identifier 59 ⟩ Used in section 55.
- ⟨ Cases like != 56 ⟩ Used in section 55.
- ⟨ Common code for CWEAVE and CTANGLE 3, 5, 6, 7, 9, 10, 12, 14, 15 ⟩ Used in section 1.
- ⟨ Compress two-symbol operator 71 ⟩ Used in section 69.
- ⟨ Copy a string or verbatim construction or numerical constant 88 ⟩ Used in section 86.
- ⟨ Copy an ASCII constant 89 ⟩ Used in section 86.
- ⟨ Expand section  $a - ^\circ 24000$ , goto restart 40 ⟩ Used in section 38.
- ⟨ Get a constant 73 ⟩ Used in section 69.
- ⟨ Get a string 74 ⟩ Used in section 69.
- ⟨ Get an identifier 72 ⟩ Used in section 69.
- ⟨ Get control code and possible section name 75 ⟩ Cited in section 92. Used in section 69.
- ⟨ If end of name or erroneous nesting, break 80 ⟩ Used in section 79.
- ⟨ If it's not there, add *cur\_section\_name* to the output file stack, or complain we're out of room 47 ⟩ Used in section 77.
- ⟨ If section is not being defined, continue 94 ⟩ Used in section 90.
- ⟨ In cases that  $a$  is a non-**char** token (*identifier*, *section\_name*, etc.), either process it and change  $a$  to a byte that should be stored, or **continue** if  $a$  should be ignored, or **goto done** if  $a$  signals the end of this replacement text 86 ⟩ Used in section 83.
- ⟨ Include files 4 ⟩ Used in section 1.
- ⟨ Initialize the output stacks 33 ⟩ Used in section 48.
- ⟨ Insert the line number into *tok\_mem* 85 ⟩ Used in sections 69, 83, and 86.
- ⟨ Insert the section number into *tok\_mem* 96 ⟩ Used in section 95.
- ⟨ Output macro definitions if appropriate 51 ⟩ Used in section 48.
- ⟨ Predeclaration of procedures 8, 11, 13, 16, 30, 35, 39, 44, 49, 53, 65, 70, 84, 91, 99, 101 ⟩ Used in section 1.
- ⟨ Private variables 20, 26, 32, 37, 42, 45, 52, 57, 62, 66, 68, 82 ⟩ Used in section 1.
- ⟨ Put section name into *section\_text* 79 ⟩ Used in section 77.
- ⟨ Read in transliteration of a character 102 ⟩ Used in section 100.
- ⟨ Scan a definition 93 ⟩ Used in section 90.
- ⟨ Scan a verbatim string 81 ⟩ Used in section 75.
- ⟨ Scan an ASCII constant 76 ⟩ Used in section 75.
- ⟨ Scan the C part of the current section 95 ⟩ Used in section 90.
- ⟨ Scan the section name and make *cur\_section\_name* point to it 77 ⟩ Used in section 75.
- ⟨ Set initial values 21, 23, 27, 46, 58, 63, 78 ⟩ Used in section 2.
- ⟨ Skip ahead until *next\_control* corresponds to @d, @<, @\_ or the like 92 ⟩ Used in section 90.
- ⟨ Typedef declarations 19, 31 ⟩ Used in section 1.
- ⟨ Update the data structure so that the replacement text is accessible 97 ⟩ Used in section 95.
- ⟨ Was an '@' missed here? 87 ⟩ Used in section 86.
- ⟨ Write all the named output files 50 ⟩ Used in section 48.