

PREPROCESADOR DE ALGOL 60 POR EL METODO AED

Por Fco. Javier Garijo y M.^a Felisa Verdejo

INTRODUCCION

Este trabajo tiene como objeto construir un preprocesador para el ALGOL 60, utilizando la técnica A.E.D.

Para ello hemos seguido los siguientes pasos:

1º).- Descripción del preprocesador.

Admitirá como inputs una cadena de caracteres terminales del ALGOL y producirá como outputs una cadena de unidades sintácticas codificadas según el código interno que les asignemos.

Para la elección de las unidades sintácticas y la clasificación de los caracteres terminales del lenguaje fuente nos hemos servido del "Rapport del lenguaje de referencia del ALGOL-60", de la carta sintáctica, así como del manual de IBM para la representación en la 70-90 de los caracteres.

Inicialmente la única condición que impusimos a las unidades sintácticas, fué que a partir de ellas un analizador sintáctico pudiera llegar hasta el axioma de la gramática, quedando la elección de una u otra a nuestro criterio.

Una vez elegidas, al intentar construir el autómata que las reconocía, vimos que era no determinista, lo cual complicaba extraordinariamente el problema. Por ello rechazamos algunas de ellas y elegimos otras de forma que el autómata fuese determinista.

Una vez determinadas las unidades sintácticas y clasificados los caracteres terminales, procedimos a su descripción según el sistema A.E.D.

2º).- Construcción del diagrama de estados a partir de la descripción del preprocesador

3º).- Paso del diagrama de estados a la construcción del autómata finito y determinista.

4º).- Programación.

El lenguaje utilizado para la implementación del autómata es el FORTRAN IV.

Un organigrama del mismo lo damos en las siguientes páginas.

1 - La función de nuestro programa es clasificar los caracteres de entrada en grupos de símbolos - unidades sintácticas - y emitir una representación suya, código, que será utilizada

por el analizador sintáctico.

Para ello necesitaremos "algo" que tome un caracter de la cadena de inputs y cuando esté clasificado, tome el siguiente. Esta función la realiza la subrutina OCS, cuya variable alfanumérica CAR contiene siempre el caracter que se analiza.

Partimos de un estado inicial S y según el contenido de CAR, pasamos a un estado determinado por la tabla de transición del autómatas, con el siguiente pasaremos a otro estado y así hasta que lleguemos al estado final.

Estos caracteres los vamos guardando en una palabra, el problema que se presenta es el de concatenar estas palabras con el nuevo caracter, contenido en CAR, esto se hace por medio de una subrutina en MAP, que se ocupa de correr el número de posiciones adecuadas al caracter en CAR, (según la posición que tenga que ocupar en la palabra donde se guarda) y de meterlo.

El número de caracteres en una palabra, viene controlado por un contador, cuando el número de caracteres excede la capacidad de la palabra, caben las siguientes opciones:

- En unidades sintácticas tales como Identificador, Caracter Compuesto, etc., para las que reservamos un máximo de 6 caracteres, la llegada del 7º provocará un mensaje de error, indicando que el número de caracteres excede de los permitidos e ignorando el resto.
- Para otras unidades sintácticas, tales como delimitador de parámetro, cadenas, etc., el 7º caracter pasa a ser el primero de una nueva palabra; otro contador controla el número de palabras permitidas, cuando se sobrepasa se emite igualmente código de error, ignorando el resto.

2 - Construcción de las tablas de símbolos y emisión de código.

Cada unidad sintáctica viene representada por una matriz con dos subíndices (i, j) el primero indica el código asignado y el segundo el lugar que esa unidad sintáctica ocupa dentro de la tabla, así por ejemplo (1, 15) = A B I Z indica que A B I Z es un identificador y ocupa el lugar 15 en la tabla de identificadores.

El código de las unidades sintácticas viene dado en el listado del programa.

Dentro de cada unidad sintáctica la tabla la construye el programa de la forma siguiente: cada vez que un grupo de caracteres son reconocidos como una U.S. el índice j se incrementa en una unidad de forma que queda posicionado para que la siguiente U.S. sea almacenada en la siguiente posición de la tabla, realizando primero una comparación con los elementos anteriores, para ver si esta U.S. está ya en la tabla, si está, el índice j no avanza. Como vemos, la construcción de la tabla y la emisión de código son operaciones simultáneas.

En el caso de U.S. tales como palabra reservada, caracter simple, caracter compuesto, ... la tabla está ya construída: una vez reconocida la U.S. para emitir el código, se

busca secuencialmente en la tabla el elemento correspondiente y se emite el código asociado.

Las funciones especiales del ALGOL tales como SIN, COS, ... las hemos tratado como identificadores y constituyen las 8 primeras posiciones de la tabla de identificadores.

Por último destacaremos que la palabra reservada COMENT lleva consigo la acción de ignorar todos los caracteres que le siguen hasta la llegada del ;

3 - Tratamiento de los errores.

Los errores detectados por el programa son los siguientes:

- Si después de comparar el caracter en CAR con los posibles caracteres del alfabeto no es ninguno de ellos se produce el error 1 - el caracter que se analiza no pertenece al alfabeto.
- Cuando el número de caracteres que forman una U.S. excede de los permitidos (6 por palabra, y un número determinado de palabras) se produce el error 2.- el número de caracteres excede de los permitidos, el resto será ignorado.
- Las unidades sintácticas que pueden empezar por una ' son tres: factor de cuadratura, palabra reservada, cadena, luego detrás de la comilla puede aparecer un + , - , dígito, letra, C, si no aparece ninguno de estos caracteres, se produce un mensaje de error - la ' es ilegal en el contexto y se vuelve a S para analizar el caracter que le sigue.
- Una palabra reservada va entre comillas, si le falta la comilla final se produce un mensaje indicando el error, pero se reconoce la palabra y se emite el código correspondiente.

Si al comparar la palabra con la tabla de palabras reservadas no coincide con ninguna de ellas, se emite un mensaje que indica que esa palabra no es una palabra reservada a pesar de aparecer entre comillas.

- La unidad sintáctica CADENA viene delimitada por ' (' al comienzo de la cadena, si aparece en la cadena de inputs ' (y el símbolo siguiente no es una ' , se da un mensaje que indica que el caracter correcto debe ser una ' , y que se toma como si lo fuera realmente, continuando el análisis de la cadena.

CONSTRUCCION DE UN PREPROCESADOR PARA EL ALGOL:

Descripción del P.P. por el sistema A E D.

```

BEGIN { <descripción símbolos> } END
BEGIN { <descripción clase de caracteres> } END
FIN

```

<descripción símbolos> :: =

```

ID (1) = letra { letra / dígito }
ESS (2) = dígito { dígito }
FC (3) = 10+ dígito { dígito } / 10 - dígito { dígito } / 10 dígito { dígito }
PD (4) = . dígito { dígito }
DP (5) = ) letra { letra } : C / ,
C (6) = '(' { letra / dígito / CECC / CEDP / NA / IGNORE } ')'
P.R (7) = 'BEGIN' / 'END' / ...
C.S (8) = + / - / * / ' / . / // / ( / ) /
C.C. (9) = # / * / * / . = / . , / ( / / ' / ' / ) /

```

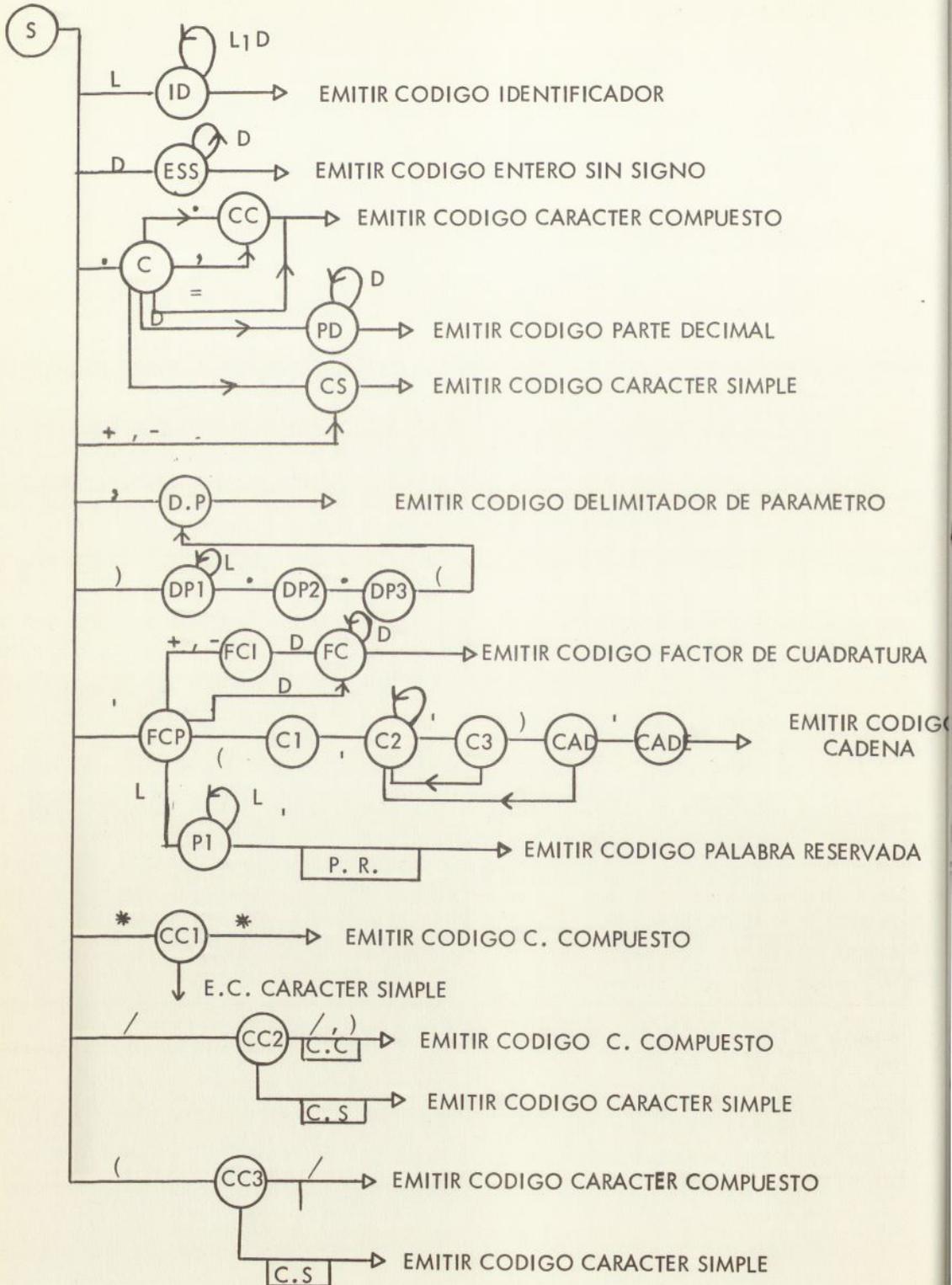
<descripción clases caracteres> :: =

```

LETRAS = / A B . . . . Z /
DIGITOS = 0 1 . . . . 9
CECC = / * . ( '
CEDP = , )
NA = + , -
IGNORE = /

```

a partir de la descripción del preprocesador, construimos el diagrama de transición siguiente:



97

78

LEER CANACT.
REGAD ALI: 131

INICIALIZAR
CONTADORES
S, SA, JB, MC, M
NA = 0

CALL DCJ

CAR: LEMBA

SI

GF GF 99

CAR: DIXITO

SI

GF GF 200

CAN: /

GF GF 301

CAN: *

GF GF 302

CAN: 0

GF GF 303

CAN: (

GF GF 304

CAR: !

GF GF 305

CAN: 3

GF GF 400

CAR:)

GF GF 401

CAN: +

GF GF 666

CAN: -

GF GF 666

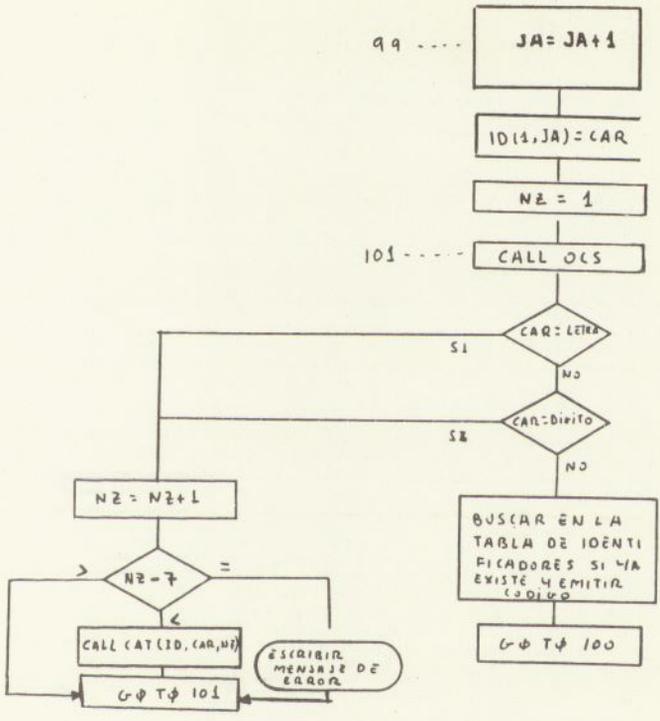
CAN: \$

GF GF 97

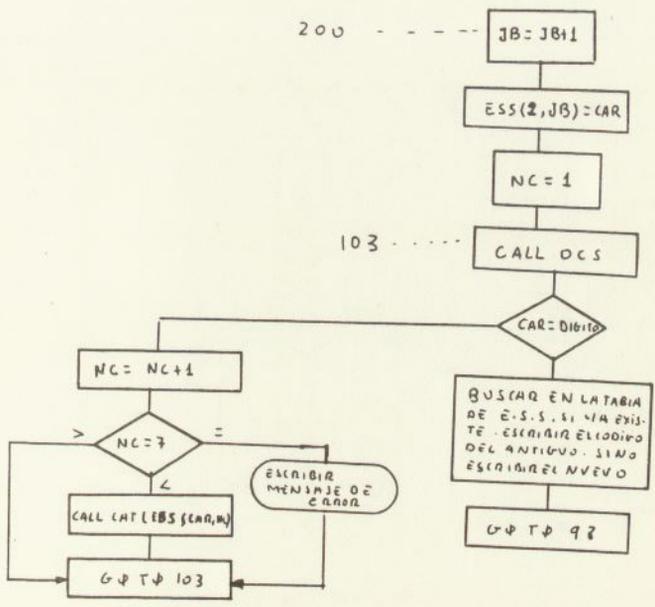
NO

ERROR 1

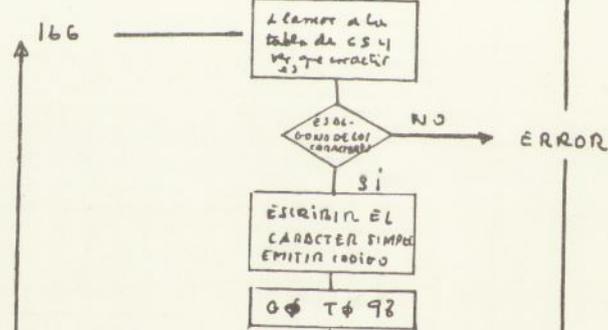
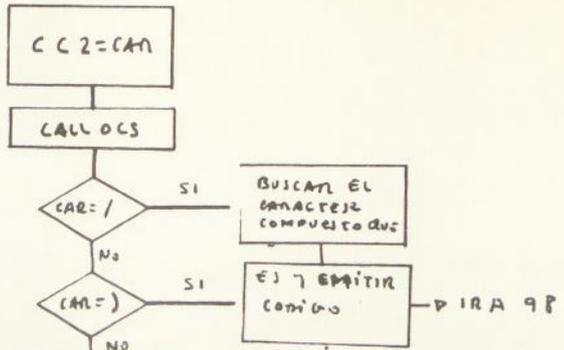
99 - Es un identificador



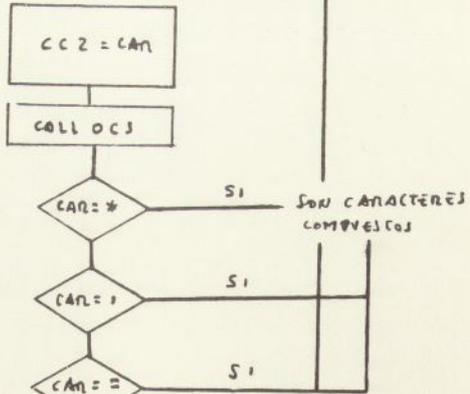
200 - Es entero sin signo



301 - Es caracter simple o compuesto

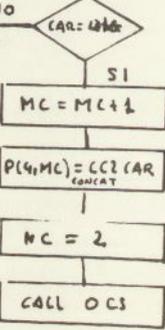


302



PARTE DECIMAL .. 310

312



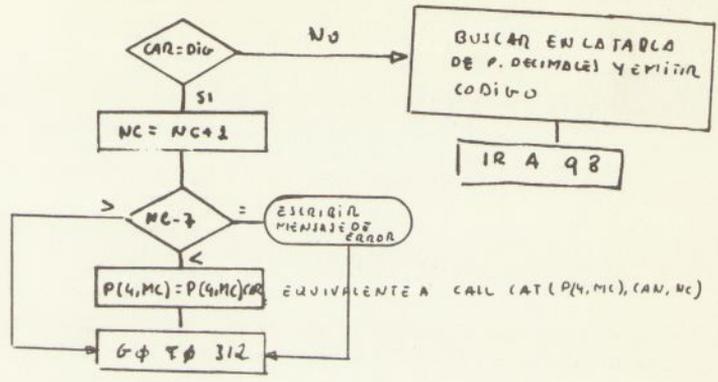
166

CONACTER SIMPLE

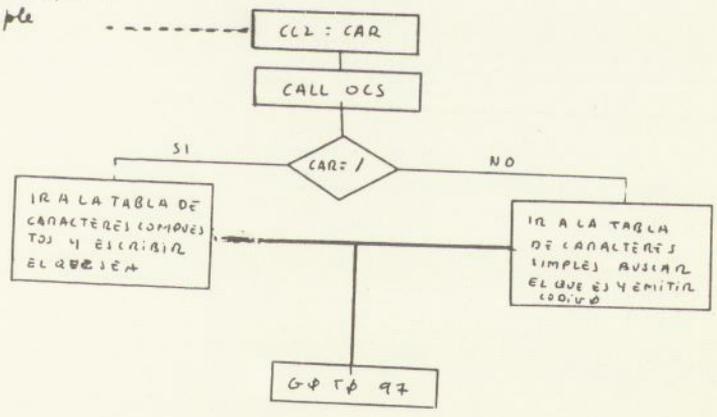
SON CARACTERES COMPLETOS

IR A 98

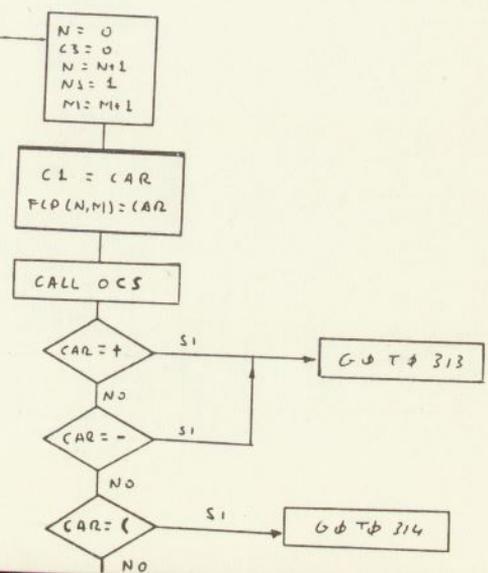
ERROR

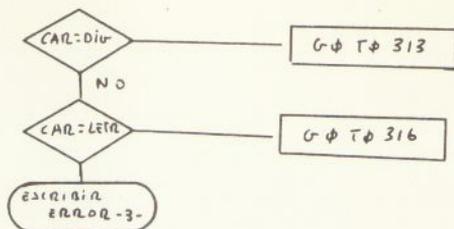


304 - Caracter compuesto o simple

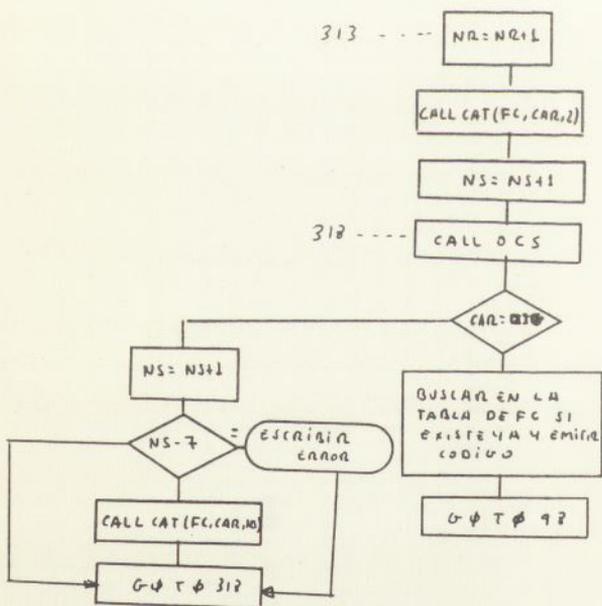


305 -

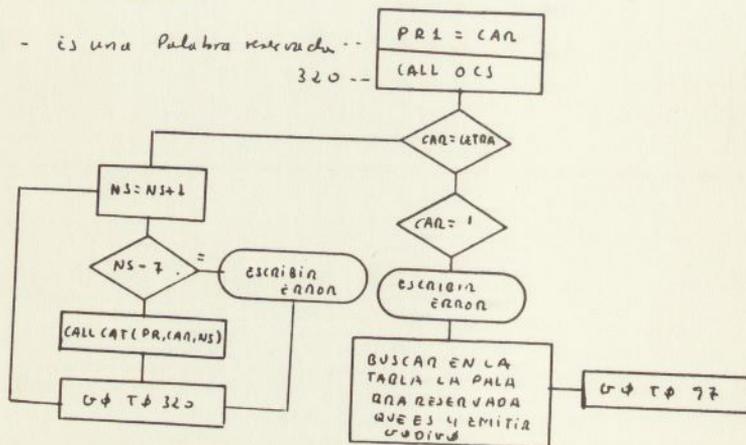


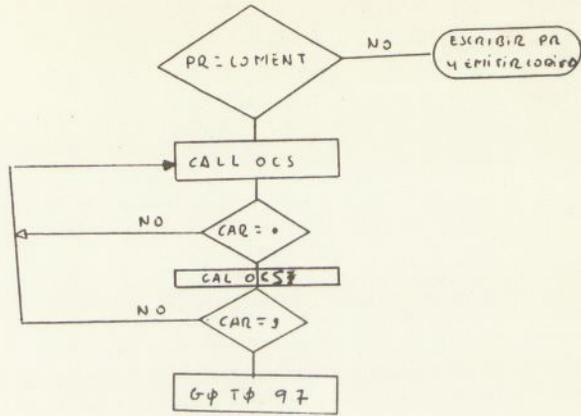


313 - Es Factor de madadura

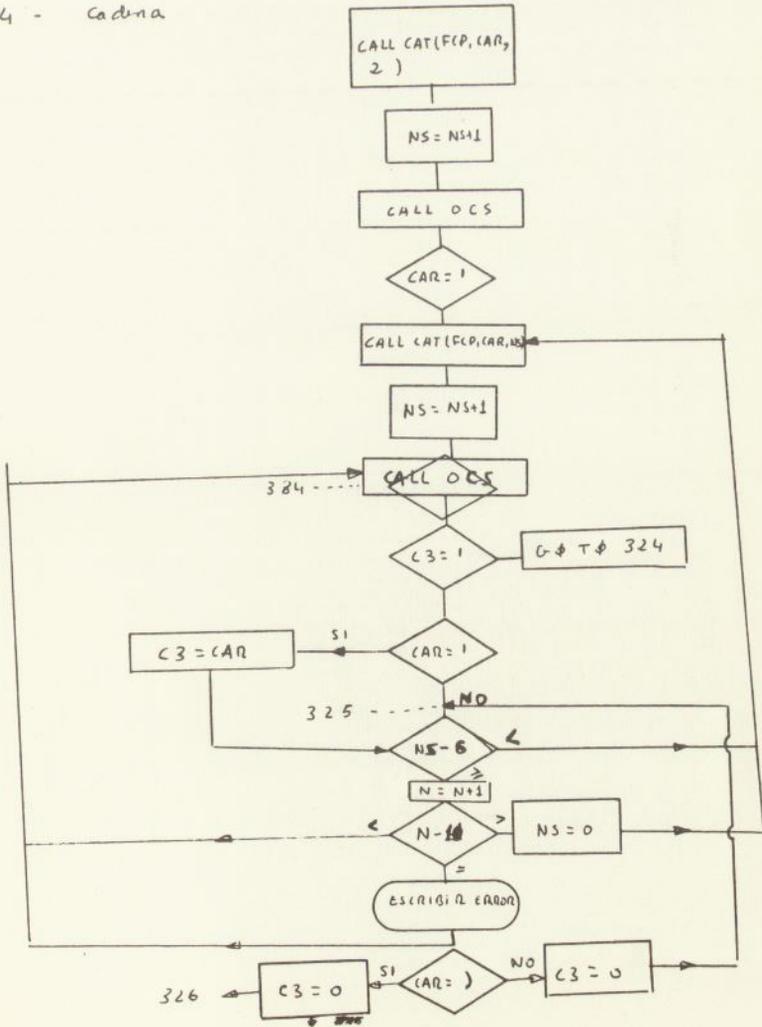


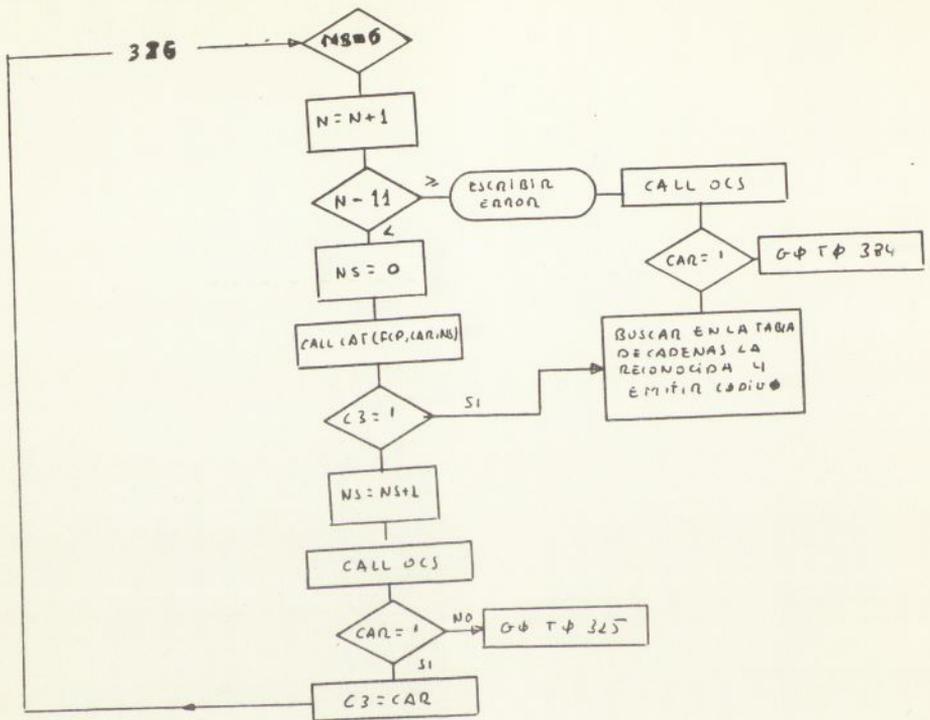
314 - es una Palabra reservada



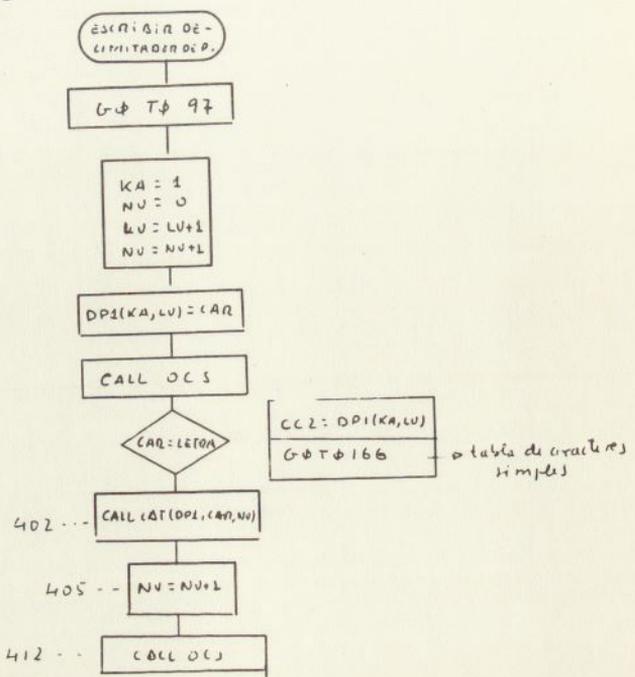


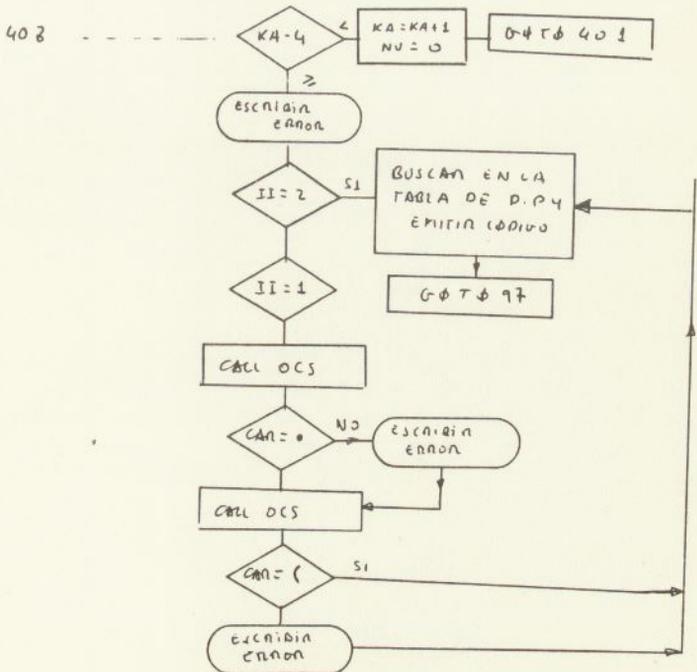
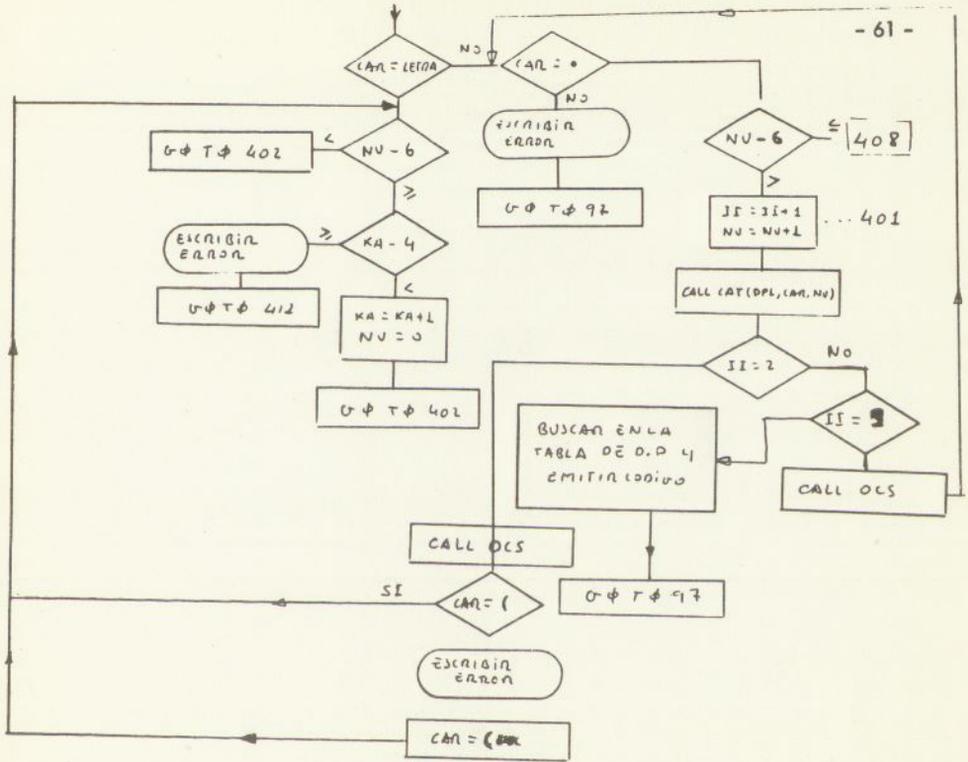
314 - Cadena





400 - delimitador de parámetro





SUBROUTINA OCS

