

## EL ANALISIS SINTACTICO AUTOMATICO COMO UNA AYUDA PARA LA ELABORACION DEL DICCIONARIO

Por: Ignacio del Campo, Isabel González, María Teresa Molina, Francisco Marcos.

En este artículo vamos a desarrollar algunos de los problemas que hemos encontrado en nuestros intentos de mecanizar el Diccionario Histórico (D.H.) de la Lengua Española. Nuestro proyecto experimental es colaboración de la Real Academia Española (R.A.E.) y del Centro de Cálculo de la Universidad de Madrid (CCUM).

Existen, verdaderamente, muchos procedimientos para desglosar las frases de un texto, y estos procedimientos son en general bastante satisfactorios. Nuestro objetivo no consiste en desglosar frases, pero, sin embargo, nos parece obvio declarar que nuestra investigación debe basarse en un buen sistema de desglose. Damos por supuesto que en nuestras frases obtenemos expresiones limitadas sintácticamente, esto es, que en nuestras frases no tenemos que tratar palabras pertenecientes a oraciones cuyos verbos no están incluidos en el texto dado en la frase. Por ahora el mejor método para entresacar las frases es limitarlas por signos de puntuación. Así, estando nuestro texto dividido por puntos o punto y comas, analizamos tantas oraciones como verbos se pueden incluir entre estos signos. Pero, y debemos advertirlo cuanto antes, todavía no estamos tratando oraciones complejas, sino que por el contrario sólo analizamos estructuras más bien simples de tipo

Sujeto                  Verbo                  Complemento Directo

lo cual significa realmente que nos encontramos con problemas de determinantes, límites debidos a inflexiones, coherencias o concordancias, etc., en lugar de tratar con problemas exclusivamente morfológicos. Manejamos también algunos puntos del análisis semántico.

Nadie espere, por consiguiente, de un trabajo tan claramente limitado ningún descubrimiento maravilloso. Nuestro sistema procesado por una máquina IBM 7090, asistida por una IBM 1401 es capaz de analizar sentencias tan simples como cada uno de los componentes del grupo de experimentos que mostramos más adelante.

Hemos escrito nuestros programas en SNOBOL, lenguaje indicado especialmente para el tratamiento de estructuras lingüísticas. Uno de los muchos problemas de nuestro compilador de SNOBOL es la imposibilidad de marcar el tiempo transcurrido en la ejecución de la tarea realizada por el ordenador, pero hemos calculado que la lectura, análisis e impresión del listado adjunto dura poco más de tres minutos.

El mensaje de error es HA FALLADO, lo cual indica que algo ha funcionado mal desde el principio. Otros mensajes de error son LA PALABRA NO ESTA EN NUESTRA TABLA, lo cual quiere decir que estamos usando una palabra inexistente en nuestro léxico (tabla 1), ESTA FRASE NO FORMA ORACION, esto es, hemos obtenido un grupo de palabras sin organización gramatical. Por organización gramatical entendemos nuestra "gramática" de la tabla 2. EL SUSTANTIVO NO CONCUERDA CON EL ARTICULO EN EL SUJETO (O EN EL COMPLEMENTO), lo cual se explica por sí

mismo; LA PALABRA NO ES ARTICULO, SUSTANTIVO NI VERBO, la palabra que estamos tratando no está en nuestra gramática, la cual sólo incluye artículos, sustantivos y verbos; NO HAY COHERENCIA ENTRE SUJETO Y VERBO, los accidentes del sujeto no son los requeridos por el verbo; LA PALABRA ESTA MAL COLOCADA, indicando que el orden de una palabra está fuera de su sitio (siempre haciendo referencia a nuestra gramática); EL SUSTANTIVO NO CONCUERDA CON EL ARTICULO, no hay equivalencia entre la forma gramatical del artículo y la del sustantivo; y NO HAY COHERENCIA ENTRE COMPLEMENTO Y VERBO, las características semánticas del complemento no son las requeridas por el verbo.

El mensaje de éxito es LA FRASE ESTA BIEN CONSTRUIDA (la oración es gramatical); si ha habido una incoherencia semántica en el análisis, pero no concuerda gramaticalmente, obtenemos A PESAR DE ELLO LA FRASE ESTA BIEN CONSTRUIDA . Con esta innovación pretendemos señalar incoherencias que sólo lo son aparentemente, como la metáfora. Por ejemplo, estableciendo que el verbo HABLAR requiere la forma + HUMANO en el sujeto, si encontramos EL PERRO HABLÓ obtendríamos el mensaje de que no hay coherencia semántica entre el sujeto y el verbo: "perro" es - HUMANO; pero, a pesar de todo, no paramos el análisis y al final obtenemos el otro mensaje de que a pesar de ello la frase está bien construída, lo que nos advierte que la frase es gramatical en un grado menor que cualquier otra totalmente coherente.

Nuestra gramática es de la siguiente manera:

S  $\implies$  SN SP

(admitimos la diferencia entre SP y SV, pero carece de utilidad en este paso de nuestro trabajo, por lo cual hacemos SP=SV)

SN  $\implies$  (DET) SUBST

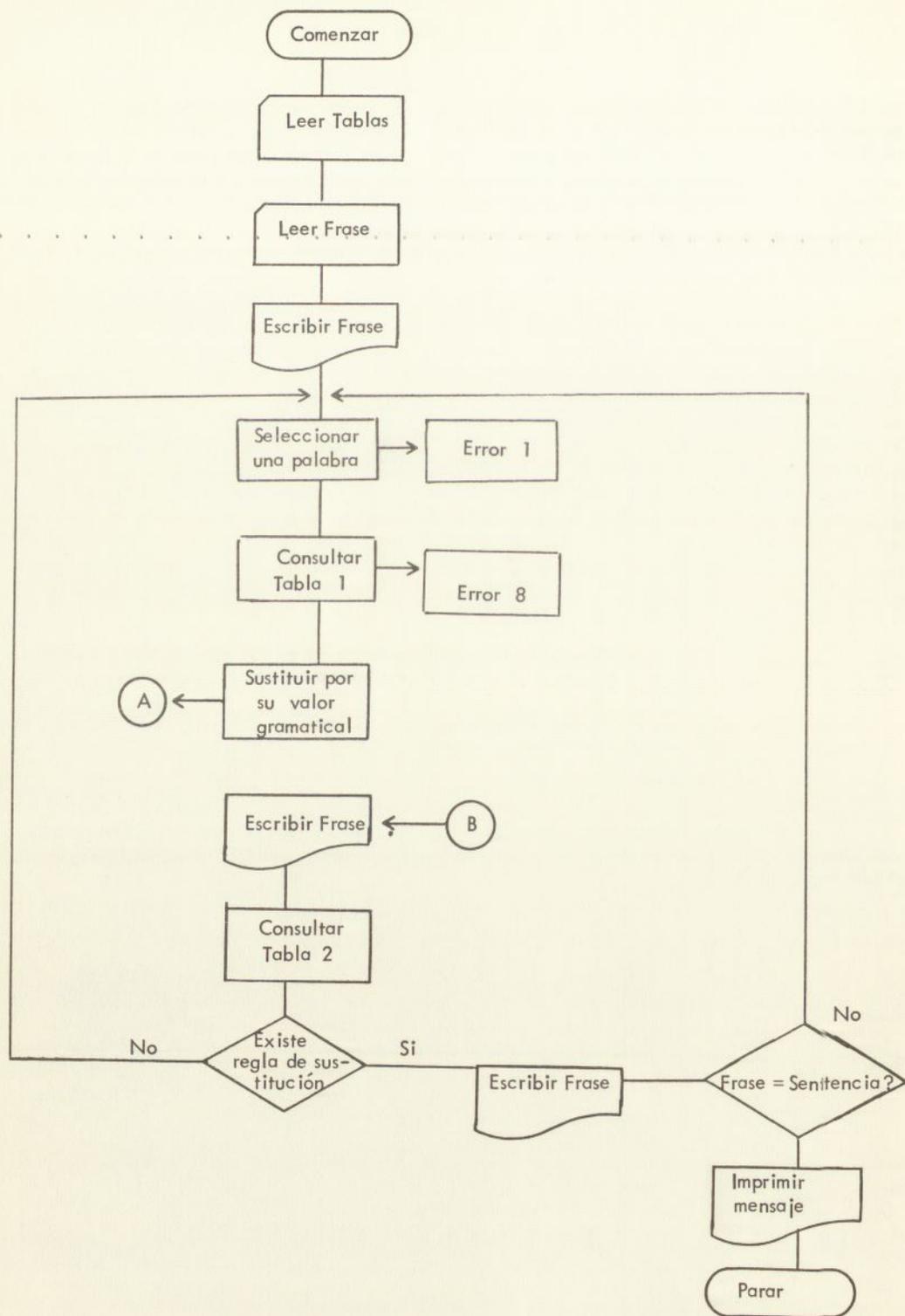
SV  $\implies$  GV [grupo verbal] (SN)

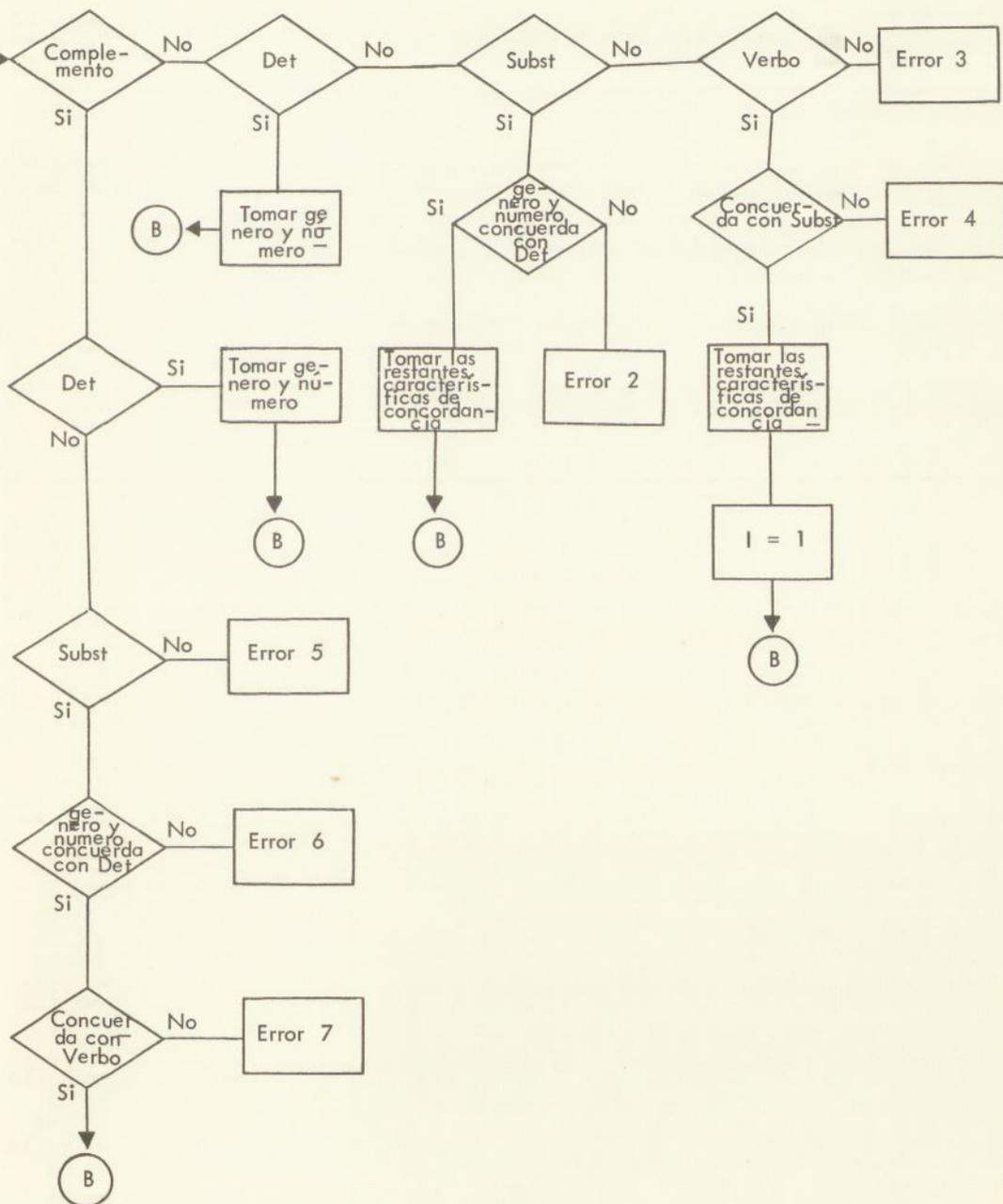
GV  $\implies$  VT - AUX

El ordenador va haciendo las sustituciones comenzando por la izquierda; si no encuentra correlación emite el pertinente mensaje de error, y si la falta de correlación es de tipo gramatical termina el proceso.



FACULTAD DE INFORMÁTICA  
BIBLIOTECA





Se incluye en el listado del programa una parte de nuestro léxico, en el cual las primeras características pertenecen al análisis gramatical, y el resto al semántico. Así PEDRO es un sustantivo, masculino, singular, en el aspecto gramatical, y animado, inteligente, en el aspecto semántico.

El verbo COME es un verbo transitivo, que requiere un sujeto animado y un complemento directo sólido. Es obvio advertir que hasta ahora nos hemos preocupado más de los problemas sintácticos que de los morfológicos. Por consiguiente operamos con verbos en tercera persona del singular. Estamos intentando construir una morfología que nos permitirá aplicar nuestro análisis a campos más amplios.

Reproducimos aquí el listado obtenido en uno de nuestros ensayos.

```

Z  [DEFINE('BLANCOS(K)', 'Z1') /{(T)
Z1 E BLANCOS =
F1 JJJ = '0'
F9 E BLANCOS = BLANCOS ' '
   JJJ = JJ + '1'
   .EEQ(JJ, '80') /F(U)
   SYSPOT = ' HA FALLADO' /{(RETURN)
U  EEQUALS(JJ, K) /F(F9)S(RETURN)
T  DDEFINE('ASTE(L)', 'V1') /{(R)
V1  ASTE =
FF11 K = '0'
FF9) ASTE = ASTE '**'
   KK = K + '1'
   .EEQ(K, '60') /F(UU)
   SSYSPOT = 'HA FALLADO' /{(RETURN)
UU  EQUALS(K, L) /F(FF9)S(RETURN)
R  DDEFINE('LINEAS(VA)', 'R1') /{(W)
R1  II = '1'
R2  SYSPOT =
   EEQUALS(II, VA) /S(RETURN)
   III = II + '1' /{(R2)
W  T TABLA =
   CUENTA = '0'
Q  SYSPIT *TEXTO/'72'*
   TTEXTO = TRIM(TEXTO)
   TTEXTO 'ENDTABLA1' /S(AE)
   TABLA1 = TABLA1 TEXTO /{(Q)
AE  SYSPOT = 'TABLA1 (LEXICO)'
   CCUENTA = CUENTA + '1'
   SSYSPOT = ASTE('15')
   CCUENTA = CUENTA + '1'
AF  TABLA1 *ENTE* '.' = /F(AG)
   SYSPOT = ENTE
   CCUENTA = CUENTA + '1'
   TTA = TA ENTE '.' /{(AF)
AG  TABLA1 = TA
   SYSPOT =
   CCUENTA = CUENTA + '1'
   SYSPOT =
   CCUENTA = CUENTA + '1'
   SYSPOT =
   CCUENTA = CUENTA + '1'
   TTABLA2 = 'DET SUST=SN.SUST=SN.VT+AUX=GV.GV SN=SV.SN SP=D.SV=SP. '
   SYSPOT = 'TABLA2 (GRAMATICA)'
   CCUENTA = CUENTA + '1'
   SSYSPOT = ASTE('18')
   CCUENTA = CUENTA + '1'
AI  TABLA2 *ANTE* '.' = /F(AA)
   CCUENTA = CUENTA + '1'
   SYSPOT = ANTE /{(AI)
AA  SYSPIT *FRASE/'72'*
   NNN = '65' - CUENTA
   SSYSPOT = LINEAS(NN)
   CUENTA = '0'
   FFRASE 'FINDETARJETAS' /S(END)
   FFRASE = TRIM(FRASE)
   FFRASE = FRASE ' '

```

```

SYSPOT = BLANCOS('48') 'ANALISIS GRAMATICAL DE LA FRASE'
CUENTA = CUENTA + '1'
SYSPOT = BLANCOS('48') ASTE('31')
CUENTA = CUENTA + '1'
SYSPOT =
CUENTA = CUENTA + '1'
I = '0'
J = '0'
K = '0'
N = SIZE(FRASE) / '2'
M = '53' - N
FRASE1 = FRASE
SYSPOT = BLANCOS(M) FRASE
CUENTA = CUENTA + '1'
SYSPOT =
CUENTA = CUENTA + '1'
J TABLA2 = 'DET SUST=SN.SUST=SN.VT+AUX=GV.GV SN=SV.SN SP=O.SV=SP.'
FRASE1 *PALABRA* ' ' = /F(E1)
TABLA1 PALABRA '=' *RESTO* ' ' /F(E8)
RESTO *CLASE* ' ' *DEMAS*
FRASE PALABRA = CLASE
.EQ(I, '1') /S(A)
CLASE 'DET' /F(B)
J = '1'
DEMAS *GENERO* ' ' *NUMERO* ' ' = /(C)
B CLASE 'SUST' /F(D)
DEMAS *GEN1* ' ' *NUM1* ' ' *GESTO*
.EQ(J, '1') /F(C)
J = '0'
EQUALS(GENERO, GEN1) /F(E2)
EQUALS(NUMERO, NUM1) /F(E2)S(C)
D CLASE 'VT+AUX' /F(E3)
SIG = DEMAS
E SIG 'SUJ' *LT* ' ' = /F(F)
GESTO LT /S(E)F(E4)
F I = '1' /F(C)
A CLASE 'DET' /F(G)
J = '1'
DEMAS *GENERO* ' ' *NUMERO* ' ' = /(C)
G CLASE 'SUST' /F(E5)
DEMAS *GEN1* ' ' *NUM1* ' ' *RESTO*
.EQ(J, '1') /F(H)
J = '0'
EQUALS(GENERO, GEN1) /F(E6)
EQUALS(NUMERO, NUM1) /F(E6)
H SIG 'COM' *LT* ' ' = /F(C)
RESTO LT /S(H)F(E7)
C N = SIZE(FRASE) / '2'
M = '53' - N
SYSPOT = BLANCOS(M) FRASE
CUENTA = CUENTA + '1'
SYSPOT =
CUENTA = CUENTA + '1'

```

```

I1 TABLA2 = 'DET SUST=SN.SUST=SN.VT+AUX=GV.GV SN=SV.SN SP=O.SV=SP. '
I TABLA2 *STR* '=' *CADENA* '.' = /F(J)
  FRASE STR = CADENA /F(I)
  N = SIZE(FRASE) / '2'
  M = '63' - N
  SYSPOT = BLANCOS(M) FRASE
  CUENTA = CUENTA + '1'
  SYSPOT =
  CUENTA = CUENTA + '1'
  EQUALS(FRASE,'O ') /S(EXITO)F(I1)
EXITO SYSPOT =
  CUENTA = CUENTA + '1'
  .EQ(K,'1') /S(AC)
  SYSPOT = '***LA FRASE ESTA BIEN CONSTRUIDA***'
  CUENTA = CUENTA + '1'
  CUENTA = CUENTA + '1'
  SYSPOT = /(AA)
E1 SYSPOT =
  CUENTA = CUENTA + '1'
  SYSPOT = '*ESTA FRASE NO FORMA ORACION*'
  CUENTA = CUENTA + '1'
  CUENTA = CUENTA + '1'
  SYSPOT = /(AA)
E2 SYSPOT =
  CUENTA = CUENTA + '1'
  SYSPOT = '*EL SUSTANTIVO NO CONCUERDA CON EL ARTICULO EN EL SUJETO*'
  CUENTA = CUENTA + '1'
  CUENTA = CUENTA + '1'
  SYSPOT = /(AA)
  CUENTA = CUENTA + '1'
E3 SYSPOT = '*LA PALABRA ' ' PALABRA ' ' 'NO ES ART,SUST, NI VER *'
  . /(AA)
E4 SYSPOT =
  CUENTA = CUENTA + '1'
  SYSPOT = '*NO HAY COHERENCIA ENTRE SUJETO Y VERBO*'
  CUENTA = CUENTA + '1'
  CUENTA = CUENTA + '1'
  SYSPOT = /(AB)
E5 SYSPOT =
  CUENTA = CUENTA + '1'
  SYSPOT = '*LA PALABRA ' PALABRA ' ESTA MAL COLOCADA*'
  CUENTA = CUENTA + '1'
  CUENTA = CUENTA + '1'
  SYSPOT = /(AA)
E6 SYSPOT =
  CUENTA = CUENTA + '1'
  A = ' EN EL COMPLEMENTO*'
  SYSPOT = '*EL SUSTANTIVO NO CONCUERDA CON EL ARTICULO ' A
  CUENTA = CUENTA + '1'
  CUENTA = CUENTA + '1'
  SYSPOT = /(AA)
E7 SYSPOT =
  CUENTA = CUENTA + '1'
  SYSPOT = '*NO HAY COHERENCIA ENTRE COMPLEMENTO Y VERBO*'
  CUENTA = CUENTA + '1'
  CUENTA = CUENTA + '1'
  SYSPOT = /(AD)

```

```

E8  SYSPOT =
    CUENTA = CUENTA + '1'
    SYSPOT = '*LA PALABRA ' PALABRA * NO ESTA EN NUESTRA TABLA*'
    CUENTA = CUENTA + '1'
    CUENTA = CUENTA + '1'
    SYSPOT = /(AA)
AB K = '1'  /(F)
AD K = '1'  /(C)
AC  SYSPOT = '*A PESAR DE ELLO LA FRASE ESTA BIEN CONSTRUIDA*'
    CUENTA = CUENTA + '1'
    CUENTA = CUENTA + '1'
    SYSPOT =  /(AA)
END  Z

```

SUCCESSFUL COMPILATION

TABLA1 (LEXICO)

\*\*\*\*\*

LAS=DET, FEM, PLU,  
 EL=DET, MASC, SING,  
 LA=DET, FEM, SING,  
 PERRO=SUST, MASC, SING, ANIMADO,  
 GATO=SUST, MASC, SING, ANIMADO,  
 PEDRO=SUST, MASC, SING, ANIMADO, INTELIGENTE,  
 PAN=SUST, MASC, SING, SOLIDO,  
 LECHE=SUST, FEM, SING, LIQUIDO,  
 FLORES=SUST, FEM, PLU, SOLIDO, NATURAL,  
 COME=VT+AUX, SUJANIMADO, COMSOLIDO,  
 BEBE=VT+AUX, SUJANIMADO, COMLIQUIDO,  
 RIEGA=VT+AUX, SUJINTELIGENTE, SUJANIMADO, COMNATURAL,

TABLA2 (GRAMATICA)

\*\*\*\*\*

DET SUST=SN  
 SUST=SN  
 VT+AUX=SV  
 GV SN=SV  
 SN SP=O  
 SV=SP

ANALISIS GRAMATICAL DE LA FRASE  
\*\*\*\*\*

PEDRO RIEGA LAS FLORES

SUST RIEGA LAS FLORES

SV RIEGA LAS FLORES

SN VT+AUX LAS FLORES

SN GV LAS FLORES

SN GV DET FLORES

SN GV DET SUST

SN GV SN

SN SV

SN SP

0

\*\*\*LA FRASE ESTA BIEN CONSTRUIDA\*\*\*

ANALISIS GRAMATICAL DE LA FRASE  
\*\*\*\*\*

EL PERRO COME LA LECHE

DET PERRO COME LA LECHE

DET SUST COME LA LECHE

SN COME LA LECHE

SN VT+AUX LA LECHE

SN GV LA LECHE

SN GV DET LECHE

\*NO HAY COHERENCIA ENTRE COMPLEMENTO Y VERBO\*

SN GV DET SUST

SN GV SN

SN SV

SN SP

0

\*A PESAR DE ELLO LA FRASE ESTA BIEN CONSTRUIDA\*

ANALISIS GRAMATICAL DE LA FRASE  
\*\*\*\*\*

PEDRO RIEGA FLORES

SUST RIEGA FLJRES

SN RIEGA FLORES

SN VT+AUX FLORES

SN GV FLORES

SN GV SUST

SV GV SN

SN SV

SN SP

0

\*\*\*LA FRASE ESTA BIEN CONSTRUIDA\*\*\*

ANALISIS GRAMATICAL DE LA FRASE  
\*\*\*\*\*

PEDRO BEBE VINO

SUST BEBE VINO

SN BEBE VINO

SN VT+AUX VINO

SN GV VINO

\*LA PALABRA VINO NO ESTA EN NUESTRA TABLA\*

ANALISIS GRAMATICAL DE LA FRASE  
\*\*\*\*\*

EL GATO BEBE LECHE

DET GATO BEBE LECHE

DET SUST BEBE LECHE

SN BEBE LECHE

SN VT+AUX LECHE

SN GV LECHE

SN GV SUST

SN GV SN

SN SV

SN SP

O

\*\*\*LA FRASE ESTA BIEN CONSTRUIDA\*\*\*

ANALISIS GRAMATICAL DE LA FRASE  
\*\*\*\*\*

PEDRO COME EL PAN

SUST COME EL PAN

SN COME EL PAN

SN VT+AUX EL PAN

SN GV EL PAN

SN GV DET PAN

SN GV DET SUST

SN GV SN

SN SV

SN SP

O

\*\*\*LA FRASE ESTA BIEN CONSTRUIDA\*\*\*



ANALISIS GRAMATICAL DE LA FRASE  
\*\*\*\*\*

EL PERRO COME PAN

DET PERRO COME PAN

DET SUST COME PAN

SN COME PAN

SN VT+AUX PAN

SN GV PAN

SN GV SUST

SN GV SN

SN SV

SN SP

0

\*\*\*LA FRASE ESTA BIEN CONSTRUIDA\*\*\*

ANALISIS GRAMATICAL DE LA FRASE  
\*\*\*\*\*

EL GATO COME EL PAN

DET GATO COME EL PAN

DET SUST COME EL PAN

SN COME EL PAN

SN VT+AUX EL PAN

SN GV EL PAN

SN GV DET PAN

SN GV DET SUST

SN GV SN

SN SV

SN SP

0

\*\*\*LA FRASE ESTA BIEN CONSTRUIDA\*\*\*

ANALISIS GRAMATICAL DE LA FRASE  
\*\*\*\*\*

EL GATO BEBE LAS LECHE

DET GATO BEBE LAS LECHE

DET SUST BEBE LAS LECHE

SN BEBE LAS LECHE

SN VT+AUX LAS LECHE

SN GV LAS LECHE

SN GV DET LECHE

\*EL SUSTANTIVO NO CONCUERDA CON EL ARTICULO EN EL COMPLEMENTO\*

ANALISIS GRAMATICAL DE LA FRASE  
\*\*\*\*\*

EL PERRO RIEGA LAS FLORES

DET PERRO RIEGA LAS FLORES

DET SUST RIEGA LAS FLORES

SN RIEGA LAS FLORES

\*NO HAY COHERENCIA ENTRE SUJETO Y VERBO\*

SN VT+AUX LAS FLORES

SN GV LAS FLORES

SN GV DET FLORES

SN GV DET SUST

SN GV SN

SN SV

SN SP

O

\*A PESAR DE ELLO LA FRASE ESTA BIEN CONSTRUIDA\*

ANALISIS GRAMATICAL DE LA FRASE  
\*\*\*\*\*

LA GATO BEBE LA LECHE

DET GATO BEBE LA LECHE

\*EL SUSTANTIVO NO CONCUERDA CON EL ARTICULO EN EL SUJETO\*

Aunque en este momento estamos trabajando en los pasos preliminares, esperamos que en un futuro próximo podremos analizar frases más complicadas, colaborando en este camino de separación de lemas por medio de la distinción, por ejemplo, entre CANTO como sustantivo y CANTO como verbo, o entre las dos posibilidades de ESPERABA como primera o tercera persona del singular.