

Una formulacion morfosintáctica (bases para la codificacion)
Francisco Milla Lozano

Proponemos un método analítico de textos susceptible de codificación computacional. Con frecuencia los trabajos realizados en este sentido ofrecen principalmente una perspectiva cuantitativa donde se contabilizan las frecuencias de tal o cual categoría gramatical aprehendida anteriormente por el ordenador.

Se consiguen con ello resultados fructuosos pero insuficientes para un comentario crítico de los textos, además de condenar al ordenador a un mero recopilador, veloz e imparcial, pero ingenuo y repetitivo también. Pensamos que la ciencia Informática ofrece hoy infinitas posibilidades a la "lengua" de los textos y por ello creemos realizable un análisis morfosintáctico computacional que defina al ordenador no como un simple "contador" de elementos pre-aprendidos sino como "estructurador" de la lengua de cualquier texto.⁽¹⁾ De esta manera el estudioso adquiere un material que suele ser fruto de agotadores e inútiles esfuerzos en la mayoría de los casos.

Antes de aplicar el proceso de análisis, el banco de datos tiene almacenadas las siguientes listas:

- a) Gramática de descomposición.
- b) Bases sustantivas del corpus.
- c) Formantes flexivos y afijos del corpus.
- d) Palabras funcionales.

(1) ALVAR EZQUERRA, M.: Proyecto de lexicografía española.
Barcelona 1976; pp.: 174-175.

Esta premeditación metodológica responde a una postura estructuralista en el análisis morfosintáctico y, principalmente, a una metodología de segmentación de la secuencia gráfica en "Constituyentes Inmediatos". Obviamente esta práctica de descripción lingüística no ofrece una herramienta teóricamente pura pero, al igual que la segmentación morfológica distingue núcleos constantes y variables finitas por su utilidad práctica, nos hemos visto limitados a su aplicación en una obra poética concreta que determina tanto los problemas metodológicos como la explicitación teórica.

1.- BASES SUSTANTIVAS.

La búsqueda de una base invariable o susceptible de invariabilidad donde asentar "L." (Lexema) no es posible en lengua castellana; sólo puede ofrecerse algún acercamiento en corpus restringidos y resulta inoperante frecuentemente. Proponemos por ello una información sustantiva que introduzca sus bases bien como ítemes funcionales susceptibles de formulaciones posteriores, o bien como complejos semánticos cerrados e ítemes léxicos abiertos, a) y b) respectivamente en los ejemplos:

a) ABANIC-: (se entiende para abanico — L.1.—
y para abanic- — L.3.—) $\left. \begin{array}{l} \text{L.1.} \\ \text{L.2.} \end{array} \right\}$ —
ALT-: (se entiende para alt- —
(alta, altura, altar...))

b) ABANICO: Nombre, común, concreto, - animado,
material, objetual, - natural, (artificio "para hacerse aire" (1)).

(1) R.A.E. (Real Academia Española): Diccionario de la lengua española. Espasa-Calpe. Madrid. 1958^{2.}

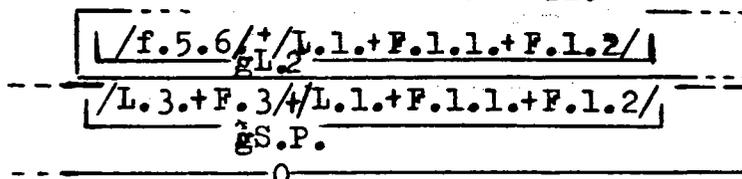
Los entrecorillados responden a una información léxica, en este caso académica, seleccionadora del significado real del morfema en su contexto oracional. Este criterio sigue el capítulo "Componentes de una teoría semántica", en la obra de Katz y Fodor, La estructura de una teoría semántica, pp. 36 y ss.

ALTO: - nombre, - verbo, estado, cualitativo,
- concreto, espaciado, dimensional, ho-
rizontal, ("superior a otro término").

Las opciones descriptivas a) y b) no son excluyentes: la primera da una información operacional que el ordenador almacena y por reglas posteriores de concurrencia en el decurso las identifica como L.2. (adjetivos), L.3. (verbos), etc.; la segunda opción posibilita una ordenación semántica básica que en nuestro caso hemos rechazado puesto que la labor del ordenador se supedita exclusivamente al reconocimiento de segmentos ya introducidos.

Sobre las bases introducidas se opera con un corpus de "reglas de desarrollo en el decurso" que describen y relacionan la ordenación morfológica, los elementos flexivos y afijales y los grupos morfológicos que pueden formar; hay por último una descripción de la ordenación oracional. Así sobre abanico incidirán:

- a) La fórmula sustantival: /[†]A.1. + L.L. [†]A.2 + F.Π/,
posibilitando: abanico, abanicos, abanicitos, abaniquitos,....?
- b) El grupo gL.1. que permite el análisis de:
el abanico, los abanicos, este abanico, abanicos blancos, estos abanicos tan blancos, etc.,.
- c) Las reglas de enunciado que reconocen secuencias como "este abanico da aire", cuya identificación se mostrará así:



A continuación se explicitan algunos de los apartados anteriores

2.- TAXONOMIA Y FORMULACION.

El análisis parte de una taxonomía convencional de los elementos operacionales que responde, a modo de referencia, a los siguientes índices:

- L.: Lexema.
 L.1.: " sustantivo.
 L.2.: " adjetivo.
- F.: Formantes (variables finitas).
 F.1.: " de sustantivo.
 F.2.: " de adjetivo.
 F.2.3.: " de adjetivo: grado...
 F.3.: " de verbo.
 F.3.1.: " de verbo: tiempos simples.
 F.3.2.: " de verbo: tiempos compuestos.
 F.3.3.: " de verbo: Persona y número...
 F.4.: " de verboides.
 F.4.1.: " de verboides: infinitivo...
 F.5.: " de pronombres.
 F.5.2.: " de pronombres indefinidos...
- G.: Transpositores.
 G.1.: " coordinantes...
- I.: Interjecciones (formas, grupos y enunciados).
- A.: Aumentos lexicales.
 A.1. " " prefijos.
 A.1.3.: " " prefijos cualitativos.
 A.2.: " " infijo-sufijales...

Cada uno de estos taxos desarrollan posteriormente sus marcas gráficas correspondientes manifestando la variabilidad de la ocurrencia en el decurso de la base que complementan:

a) F.1.1.a. Formantes de sustantivo: género masc.:

$$\text{MARCAS: } \left| \begin{array}{l} -o \\ -e \\ -or \\ -\emptyset \end{array} \right\} + F.2 \left| \begin{array}{l} \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \end{array} \right. \left(\frac{\text{---}}{\text{---}} = \sqrt{\text{A.1} + \text{L}^{\pm}} \right)$$

b) F.5.1.1. Formantes de pronombres personales.

Primera persona:

$$\text{MARCAS: } \left| \begin{array}{l} yo \\ nos \\ nosotr- \\ \end{array} \right\} + F.1.$$

c) A.1.1. Prefijos semiautónomos:

MARCAS:

$$\left. \begin{array}{l} /aero- (1)/ \\ /ambi- / \\ /antropo- / \\ /apo- / \\ /aristo- / \dots \end{array} \right\}.$$

Una vez definidos los taxos a través de sus fórmulas de entrada se delimitan estas en sus correspondientes fórmulas "desarrolladas" o restricción por la ocurrencia constante en el ~~curso~~. Así la fórmula de desarrollada u ordenación constante de F.5. será:

$$/F.5. \dagger F.2.3.1.a. \dagger F.1.1. \dagger F.1.2./.$$

Cuando los elementos hasta ahora descritos pueden formar grupo entre sí y aparecen en el decurso con cierta frecuencia los formulamos como "grupos" o conjuntos de elementos morfológicos. De esta forma el taxo F.5.7. (formantes de pronombres artículos) confecciona la fórmula constante del grupo:

$$/F.5.7. \dagger F.1./ \dagger / \dagger A \left\{ \begin{array}{l} L. \\ F.5 \end{array} \right\} \dagger A \dagger F.2.3.1.a \dagger F.1./.$$

2.1.- Bases Grupales:

La conformación de grupos sintagmáticos se elabora sobre las delimitaciones morfológicas anteriores realizando una segmentación desde la secuencia gráfica y sin acceder en ningún caso a sus posibles funciones. Los grupos se organizan como ordenamientos morfológicos complejos cuyos núcleos se han identificado anteriormente como sustantivos, adjetivos, etc.; de esta identifica-

(1) El guión de las marcas de esta grupo da entrada a las bases sustantiva y adjetiva bajo la fórmula restringida: $— = +L \dagger A.2 \dagger \left\{ \begin{array}{l} F.1 \\ F.2 \end{array} \right\} |$.

ción reciben, por ello, las denominaciones de:

- gL.1.: Grupo sustantival.
- gL.2.: " adjetival.
- gL.3.: " verbal.
- gL.4.: " adverbial.
- gG.3.: " preposicional.
- gL.3.: " no verbal...

Estas bases grupales desarrollan después sus fórmulas de ocurrencia:

$$a) \text{ gL.1.} = \boxed{+/D/+ /L.1./+ /M/} \quad (1).$$

$$b) \text{ gL.3.} = 1)$$

$$\boxed{+/F.3.4/+ \left\{ \begin{array}{l} /F.3.6- \\ /"est"- \\ /"parec"- \end{array} \right\} + \left\{ \begin{array}{l} -F.3/ \\ -F.4/ \end{array} \right\} + /N/+ /L.3+ F.4/} \quad (2)$$

2)

$$\boxed{+/ \left\{ \begin{array}{l} /F.3.4/ \\ /F.3.6/ \end{array} \right\} + /L. \left(\begin{array}{l} \gamma F.3.6- \\ \gamma "est"- \\ \gamma "parec"- \end{array} \right) + \left\{ \begin{array}{l} F.3 \\ F.4 \end{array} \right\} /N/+ /L.3+ F.4/}$$

2.1)

$$\boxed{\left\{ \begin{array}{l} /"and"+F.3/ \\ /"sol"+F.3/ \\ /"ten"+F.3/ \\ /"ven"+F.3/ \\ /F.3.7.12/ \end{array} \right\} + /L.3 \quad F.4/}$$

$$(1) /D/ = \left(\begin{array}{l} \text{gL.2/} \begin{matrix} n \\ + \\ n \end{matrix} \text{ /gG.3/} \begin{matrix} n \\ + \\ n \end{matrix} \\ /F.5.7/+ \text{ /gG.3/} \\ \left\{ \begin{array}{l} /F.5.2/n \\ /F.5.3/n \\ /F.5.4/n \\ /F.5.5/n \\ /F.5.6/n \end{array} \right\} \end{array} \right) \quad /M/ = \left(\begin{array}{l} \text{gL.2/} \begin{matrix} n \\ + \\ n \end{matrix} \\ \text{gL.1/} \begin{matrix} n \\ + \\ n \end{matrix} \\ \text{gG.3/} \begin{matrix} n \\ + \\ n \end{matrix} \\ [0] \end{array} \right).$$

(2) Para formar gL.3 deben de darse los complejos señalados o también:

$$\boxed{/L.3+ \left\{ \begin{array}{l} F.3 \\ F.4 \end{array} \right\} \left\{ \begin{array}{l} / \\ / \\ n \\ / \\ n \end{array} \right\}}$$

2.2.- Segmentación sintagmática:

El análisis sintagmático que posibilitan las fórmulas grupales posee la utilidad de funcionar como operador coadyuvante a la segmentación gráfica y el análisis morféxico determinando el volumen textual, la parca del decurso donde aplicar las reglas y fórmulas precedentes. De esta manera el enunciado responde al signo:

$$"0" = \boxed{\pm / -gL.3 / \pm / gL.3 / \pm / -gL.3 /}$$

Puede observarse que la presencia de /gL.3/ para determinar la presencia de un enunciado es arbitraria. La especificación de la formulación del enunciado se acompaña de un listado de variantes que extenderían abusivamente este apartado; en todo caso, se delimitan como enunciados todas aquellas secuencias grupales o morfológicas seccionadas para el análisis como incluidas en "0": [], posean o no L.3 .

Desde esta base formulamos igualmente enunciados complejos:

$$\boxed{\pm / G.1 / + 0} + \boxed{/ G.1 / + 0}^n = \text{Coordinante.}$$

$$\boxed{0} + \boxed{/ G.2 / + 0}^n = \text{Subordinante primario ("0.1").}$$

$$\boxed{0} + \left[\begin{array}{l} / G.1 / \\ / G.2 / \end{array} \right] + \boxed{/ -gL.3 /}^n = \text{Enunciado subordinante complejo ("0.L.2").}$$

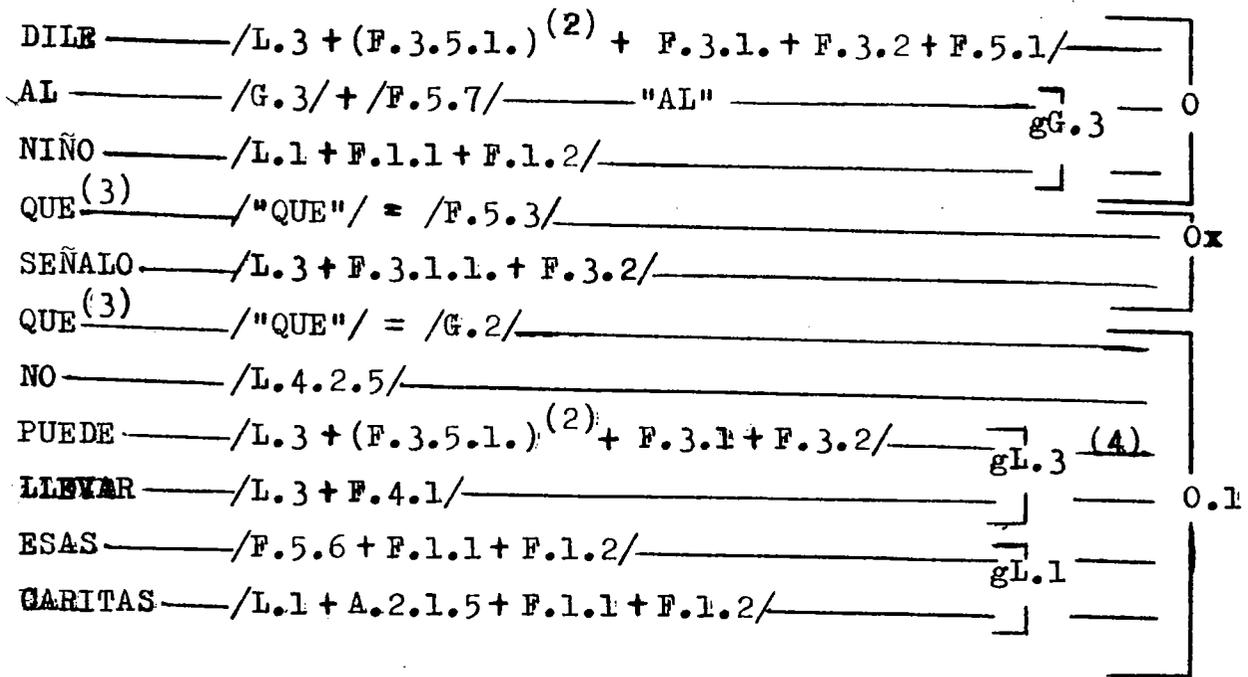
$$\boxed{0.1} + \boxed{/ G.2 / + 0}^n = \text{Subordinante secundario ("0.2")}$$

3.- PROCESAMIENTO.

Dadas las bases sustantivas, la secuencia textual, las fórmulas de entrada y de desarrollo, los grupos morfológicos, la segmentación sintagmática y las fórmulas funcionales, el ordenador realizará un análisis como el

que ejemplifica las bases y su operatividad:

"DILE AL NIÑO QUE SEÑALO QUE NO PUEDE LLEVAR ESAS CARITAS".(1)



NIVEL
de
análisis

/PALABRA/

/GRUPO/

ENUNCIADO

(1): Hemos incluido el vocablo "caritas" intencionalmente ya que queremos subrayar que el análisis se efectúa sobre textos seleccionados previamente y en los que, por supuesto, no caben enunciados ambiguos o agramaticales.

(2): El paréntesis especifica la presencia de variante lexemática en los verbos "decir" y "poder". El programa incluye un apartado que formula las variaciones y aumentos verbales y nominales.

(3): No hemos querido omitir el problema de la distinción entre "que" relativo y "que" subordinante no relativo. En algunos procedimientos sobre lenguas naturales

esta distinción, al igual que el análisis de pasivas, etc., se hace imposible (TENANT, H.: Natural Language Processing. New York. 1981). En nuestro programa se incluye una descripción de contornos textuales de "que" permitiendo enfrentar esta distinción con un margen de error mínimo en castellano. En el ejemplo que exponemos quedaría solventado mediante la orden "CH" que asigna la relativización al espacio -gL.3 inmediato.

(4): El grupo [gL.3] contiene toda secuencia verbal compleja incapaz de conformar más de un enunciado.

Sobre este procedimiento podemos exigir información cuantitativa y cualitativa requiriendo, por ejemplo, frecuencia de cada formante o de los que nos interesen; elementos y estructuras morfológicas, lexemáticas; evolución léxica y semántica de la obra si ha sido previamente ordenada; etc.

Muchas de estas peticiones, como es obvio, dependen de la preelaboración textual objeto del análisis que será útil para localizar vocablos, estructuras o sentidos de los que el autor va prescindiendo al correr de su obra. En nuestro caso, en la obra de Blas de Otero que ha suscitado este método, la advocación de "Señor" con el significado de Dios desaparece del texto del autor en un momento dado, a ello acompaña el cambio de la grafía de "Dios" a "dios" y el paso a actitudes métricas y temáticas diferentes.

Francisco MILLA LOZANO