

# EL MODELO MORFOLÓGICO FORMAL PARA SUSTANTIVOS Y ADJETIVOS EN EL ESPAÑOL

Igor A. Bolshakov  
Centro de Investigación en Computación I.P.N.  
Unidad Profesional Adolfo López Mateos  
México, D.F. 07738, México  
[e-mail: igor@pollux.cenac.ipn.mx](mailto:igor@pollux.cenac.ipn.mx)

## RESUMEN

Se propone un conjunto de reglas puramente formales para la transformación de un par {*un nombre de un lexema, sus características morfológicas*} a la forma de palabra correspondiente, donde un lexema (un conjunto de formas de palabra) es un sustantivo o adjetivo español arbitrario, y las características son el número y el género gramaticales. Entre los adjetivos se encuentran los numerales ordinales.

Se presenta una clasificación morfológica completa de los sustantivos y adjetivos del español, y se muestran las peculiaridades implicadas por las alteraciones ortográficas y la necesidad de señalar correctamente las vocales acentuadas.

Se propone también un segundo variante del modelo, en el cual la cantidad de reglas es dos veces menor, pero el volumen de información almacenada es mayor.

Ambos variantes pueden usarse para el análisis y la síntesis morfológicos de palabras del español en los sistemas de procesamiento de texto. El primer variante es preferible para la síntesis, y el segundo para el análisis.

El conjunto de reglas propuesto es un fragmento del modelo formal de la morfología total del español.

## PALABRAS CLAVES

Modelo morfológico formal, español, lexema, forma de palabra, reglas, síntesis, análisis, sustantivos, adjetivos, acento, sílabas, alteraciones ortográficas, supletivismo.

## INTRODUCCIÓN

En los idiomas romances como el español, el italiano o el francés, no existe la declinación de nombres sustantivos y adjetivos, y por eso la morfología de los nombres se considera bastante simple. Sin embargo, la descripción formal de esta morfología en el español, en plenitud total y con posibilidad de realización por una computadora, no es conocida en la literatura.

Este artículo tiene la meta de crear un modelo formal para la morfología de nombres del español. Un hecho de que este problema es todavía actual, se manifiesta particularmente en unas faltas que se encuentran en las partes morfológicas de procesadores de palabras modernos (por ejemplo, la herramienta Ortografía en Word para Windows-95, opción español [1]) y de traductores del español al inglés y viceversa (Power Translator Professional, [2]).

El objetivo del modelo morfológico formal es describir la transformación de un par {*un lexema, sus características morfológicas*} hacia la forma de palabra correspondiente. El lexema se define en la lingüística general como un nombre común (o un número del artículo correspondiente en un diccionario) de un conjunto de todas las formas de palabras correspondientes. Las características morfológicas se definen como un conjunto de esos índices gramaticales, que concretan una forma de palabra, por seleccionar entre todas las posibles para este lexema. En este caso, el lexema puede ser un sustantivo o adjetivo arbitrario, y el conjunto de las características se reduce al número gramatical para sustantivos y a una combinación de un número y un género para adjetivos.

En todas sus propiedades morfológicas, los numerales ordinales son semejantes a los adjetivos, y así se incluyen en este modelo.

Se percibe fácilmente que para adjetivos no se tienen grados de comparación. Esto se hace con premeditación, porque los grados se expresan en español por dos medios diferentes, usando flexiones o construcciones analíticas. Las últimas, para adjetivos y especialmente para verbos, son dignas de un trato separado. No se quiere analizar así las formas apocopadas que existen para pocos adjetivos y numerales ordinales: gran, buen, mal, primer, tercer, etc.

Postergando las definiciones y análisis más exactos, tendríamos algunos ejemplos demostrativos de la transformación morfológica:

{REGIÓN, (plur)} => regiones

{LÁPIZ, (plur)} => lápices

{CIENTÍFICO, (fem, plur)} => científicas

{FELIZ, (masc, plur)} => felices

{INGLÉS, (fem, sing)} => inglesa

{SÉPTIMO, (fem, sing)} => séptima

La creación del modelo formal de morfología del español comenzó desde hace mucho tiempo. En 1974 fué publicado en ruso por Igor Mel'cuk [3] el modelo formal del verbo del español. Esta obra importante fue el prototipo para investigaciones en el dominio de modelos morfológicos para muchos otros idiomas flexivos, con morfología más difícil que en español, tal como el ruso [4] o el latín [5]. El enfoque de Kimmo Koskeniemi en la morfología de lenguajes flexivos [6], que apareció posteriormente en la literatura, es menos favorable para el español.

Pero para el tema de este artículo, la obra citada de Mel'cuk y su variante más moderna publicada en español [7] no es aplicable directamente.

En primer lugar, el trabajo de Mel'cuk [3] investiga el verbo, y sólo unos de los rasgos de este modelo ligados a alteraciones ortográficas se transportan hacia adjetivos y sustantivos.

En segundo lugar, en el modelo de Mel'cuk la representación fonológica se toma como uno de los niveles básicos. Esta representación usa signos especiales de escritura fonética. Para la traducción de estos signos hacia la forma de texto habitual se usa el conjunto complejo de reglas del modelo, igualmente aplicable a palabras de todas las partes de la oración.

La complicación de estas reglas puede imaginarse con estos hechos: que en concordancia con la pronunciación moderna, el modelo tiene un signo común para las letras (*b*, *v*), un signo a veces común para los pares de letras (*i*, *y*) y (*c*, *k*) y no tiene ningún signo para la hache. Como resultado, las reglas de transformación hacia la escritura habitual reciben apariencia del diccionario, que opera no sólo con signos diferentes, sino con muchas palabras ("excepciones") en su totalidad.

Otra circunstancia, que complica considerablemente la utili-

zación de la representación fonética, es el hecho de que no puede ser igual para todos los dialectos diferentes del español. Por ejemplo, las letras *z* y *ll*, en todos los casos, y letras *c* y *d*, en algunos casos, no suenan en América Latina como en España. En otras palabras, la representación textual une los dialectos del español, mientras que la fonética los separa.

En fin, la representación fonética no es práctica, donde se investiga el procesamiento del texto, no del habla. Se pone como condición indispensable la posibilidad de traducir fácilmente del modelo morfológico formal hacia algoritmos de síntesis y sucesivamente de análisis morfológicos de palabras del español. Si algún nivel de representación interno en el modelo sólo dificulta la algoritmización, este nivel puede eliminarse.

Una diferencia más de este enfoque, en comparación con el de Mel'cuk, es la reunión de niveles literal y soprasedgmental de descripción de palabras del español. Es conocido que la acentuación, como un rasgo soprasedgmental, influye a las señales correctas de las vocales acentuadas. Las reglas de Mel'cuk se dan independientemente para los dos niveles y esto dificulta el tránsito a los algoritmos. Se usan unos signos auxiliares que ayudan a elegir el lugar del acento y después desaparecen.

El plan de exposición es como sigue. Al principio, se darán algunos rasgos generales teóricos de modelos morfológicos formales, en comparación a las gramáticas formales de N. Chomsky. Después se introducirá la clasificación morfológica precisa para adjetivos y sustantivos del español y se darán herramientas formales para la definición correcta de señales de acento y para alteraciones ortográficas. Esto permitirá dar precisamente las reglas de este modelo y la cantidad suficiente de ejemplos de funcionamiento del modelo para verificar los resultados del trabajo. Como otro variante del modelo, al final se dará un conjunto de reglas que pueden usar dos opciones de la base de un lexema en vez de la única, y la comparación de los variantes desde el punto de vista de la computación.

## MODELOS MORFOLÓGICOS FORMALES

Antes de pasar a los modelos morfológicos formales, recordaremos algunos rasgos de gramáticas generativas (GG) por N. Chomsky. Estas gramáticas se definen como un cuádruple {*CN*, *CT*, *SI*, *RS*}, donde *CN* es un conjunto de símbolos no terminales, *CT* es un conjunto de símbolos terminales, *SI* es un símbolo inicial de los no terminales, y *RS* es un conjunto de reglas de sustitución (reescritura). Las reglas de sustitución en el caso general tienen apariencia

$$S1 \Rightarrow S2 \quad (1)$$

donde las partes *S1* y *S2* son cadenas de símbolos terminales y no terminales.

Una GG trabaja como sigue: entre las reglas *RS* se toma una con la parte izquierda igual a *SI*, y *SI* se reemplaza con la parte derecha de esta regla. Por lo tanto, se inicia la formación de una cadena de derivación. Después se busca una regla, cuya parte izquierda esté contenida en la cadena indicada; en el caso de encontrarse, la parte derecha reemplaza a la izquierda, etc. El proceso de búsqueda y de sustitución se continúa hasta que toda la cadena formada contenga solo símbolos terminales.

Los símbolos terminales y no terminales pueden tener interpretaciones totalmente diferentes, según el dominio seleccionado. Son ampliamente conocidas las aplicaciones de las GG a lenguajes de programación, a varios niveles de lenguajes naturales (habitualmente a la sintaxis), a teoría de conocimiento de patrones visuales, etc.

En la aplicación directa de las GG a la morfología surge cierto obstáculo conceptual. La palabra, en su apariencia como una cadena de letras, se genera desde un símbolo inicial universal. En el proceso de la generación se seleccionan de cierto modo un lexema y unas características que transmiten un carácter concreto al resultado. La concretización se deriva desde cero, por medio de la selección arbitraria de alguna regla.

Aunque el lenguaje, en opinión de muchos lingüistas, es un transformador de un sentido a un texto y viceversa, en los límites de este transformador, según el modelo «Sentido  $\Leftrightarrow$  Texto» por Mel'cuk [8], existe la etapa morfológica que sobreentiende la transformación durante la síntesis de cada palabra aislada

$$(L, C) \Rightarrow W \quad (2)$$

donde *L* es el lexema, *C* representa las características morfológicas, y *W* es la apariencia literal de la forma de palabra sintetizada.

Como tentativa de unir la concepción de las GG con la formación de la palabra (2) llegamos a la idea de derivación dirigida con reglas, que incluye en comparación a la fórmula (1), una condición (a la izquierda del signo  $\vdash$ ):

$$L, C \vdash SI \Rightarrow S2 \quad (3)$$

Es importante notar que para cada regla concreta (3), son relevantes sólo algunas componentes de *C*, de *L* o de los índices conectados a *L*.

Una regla de sustitución debe descubrir un símbolo inicial como la cadena de morfemas *MB + MF*, donde *MB* es el símbolo del morfema de un tema radical (una base) y *MF* es el de una flexión. Esta estructura bipartita se observa para todos los sustantivos y adjetivos del español, y *MF* depende de la parte de la oración, porque la regla correspondiente tiene dos variantes.

Como otro ejemplo, una regla de sustitución reemplaza el símbolo de base *MB* a una cadena de letras, que depende de *L*. Para

ser más preciso, para *L*, como para una dirección dentro de un diccionario de bases, se extrae su base y se reemplaza el símbolo *MB*.

Como tercer ejemplo, la sustitución del símbolo de flexión *MF* por letras concretas depende, de un lado, de la clase morfológica de *L* (esta clase es un elemento de datos conectados a *L* en el diccionario), y, de otro lado, de las características *C*.

Debido a lo anterior, las reglas de un modelo morfológico formal no se acaban. Cerca de la juntura de los dos morfemas pueden tener lugar algunas variaciones ortográficas, como  $z \Rightarrow c$  (por ejemplo, LÁPIZ),  $o \Rightarrow o$  (por ejemplo, REGIÓN),  $o \Rightarrow o$  (JOVEN), etc. En estos casos, las reglas de sustitución no dependen de *L* o *C* o cantidades conectadas, pero pueden ser dependientes del contexto dentro de la cadena, es decir, contienen sustituciones de símbolos separados dependientes de símbolos adyacentes o sustituciones de pares de símbolos juntos.

## LA CLASIFICACIÓN MORFOLÓGICA DE NOMBRES Y EL PROBLEMA DE LA ACENTUACIÓN CORRECTA

Los sustantivos del español pueden dividirse en tres clases:

- clase 0: son invariables (como LUNES, RASCACIELOS, etc.);
- clase 1: agregan -s a una base en plural (como CHICO, CARA, PIE, TRIBU, etc.);
- clase 2: agregan -es a una base en plural (como VEZ, MES, PAPEL, TRACTOR, LEY, CIUDAD, etc.).

Los adjetivos del español pueden dividirse en seis clases:

- clase 0: son invariables (ISÓSCELES);
- clase 1: tienen flexiones  $\{-o, -a, -os, -as\}$  para masc. sing., fem. sing., masc. plur. y fem. plur. respectivamente (como BLANCO, FÉRREO, CARO, CIENTÍFICO). Notamos que para esta clase, la base es la forma diccionaria de masc. sing. menos la última letra -o (*blanc* para BLANCO);
- clase 2: tienen flexiones  $\{-, -, -s, -s\}$ , es decir, no varían por el género y agregan -s en plural (como GRANDE, SIMPLE, IDEALISTA, AGRÍCOLA, etc.) Entre el conjunto de las flexiones se presenta la flexión vacía «-»;
- clase 3: tienen flexiones  $\{-, -, es, -es\}$ , es decir, agregan -es en plural (como COMÚN, DÉBIL, JOVEN, BALADÍ);
- clase 4: tienen flexiones  $\{-, -a, -es, -as\}$  (como FRANCÉS, ALEMÁN, ANDALUZ, RECEPTOR);
- clase 5: tienen flexiones  $\{-e, -a, -es, -as\}$  (GRANDOTE, REGORDETE).

En la clase 2 de sustantivos y en las clases 3 y 4 de adjetivos el número de sílabas en la palabra cambia en concordancia del número o género. Mientras el lugar del acento dentro de una

base habitualmente se mantiene. Esto significa, que si en la forma diccionaria el acento cae en la sílaba última y la vocal tiene acento ortográfico, entonces después del prolongamiento de la palabra, la señal del acento debe desaparecer (para la máquina, ésta es una letra diferente: *común* vs. *comunes*, *alemán* vs. *alemanes*, *alemana*, *alemanas*).

En otros casos, dentro de las mismas clases de sustantivos y adjetivos, si el acento en la forma diccionaria cae en la segunda sílaba desde el fin y no se marca, después del prolongamiento la señal del acento debe aparecer (*joven* vs. *jóvenes*).

Se propone, que en las formas diccionarias en las clases indicadas las dos sílabas últimas no tengan diptongos, y el acento potencial pueda ser sólo en estas sílabas. Si se buscaran tales nombres de las clases en las cuales estas condiciones no se realizaran, el modelo debería complicarse un poco.

Las reglas del modelo formal deben tener en cuenta estas variaciones. Pero introducir directamente a las reglas un contador de sílabas no es particularmente natural. Ni la desmembración a las sílabas, ni tampoco las operaciones aritméticas se expresan en modo simple por reglas de sustitución. Es mejor introducir tales signos y operaciones sobre ellas, que reduzcan el problema a reglas habituales para las GG. Éstas se propondrán más adelante.

## COMPONENTES DE LAS REGLAS FORMALES

Las reglas morfológicas formales para los nombres del español se sirven de los valores siguientes.

El lexema *L* se describe en el diccionario por:

- Un morfo concreto *mt* de la base. Esta es una cadena de letras del alfabeto español. En la primera posición dentro de la cadena puede usarse una minúscula o mayúscula, en posiciones siguientes pueden usarse minúsculas o signos no terminales, que son vocales con acento potencial o consonantes variables en concordancia con la ortografía (véase abajo);
- Un índice de la parte de la oración  $Po = \{adj \mid sus \mid \}$ , donde *adj* y *sus* significan adjetivo y sustantivo respectivamente;
- Una clase morfológica *Cl* con significaciones

$$Cl = \{0 \mid 1 \mid 2\} - \text{para sustantivos y}$$

$$Cl = \{0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5\} - \text{para adjetivos.}$$

Aquí y en adelante la vertical separa alternativas posibles.

Para ser más precisos, existe también un elemento de la descripción diccionaria de los sustantivos, la señal de permisibilidad de ambos números, sólo del plural (sustantivos *pluralia tantum* como TIJERAS) o sólo del singular (*singularia tantum* como

TRÁFICO). Suponemos que esta señal se tiene en cuenta de algún modo antes de la síntesis morfológica, en la etapa de composición del par  $\{L, C\}$ . En nuestro modelo esta señal no figura.

El componente único de las características morfológicas para los sustantivos es su número gramático: Núm =  $\{sing \mid plur\}$ .

Las características de los adjetivos contienen dos componentes, del número Núm =  $\{sing \mid plur\}$  y del género Gen =  $\{masc \mid fem\}$ . Es conocido que el significado neutral del género puede pertenecer en español sólo a los sustantivos, como un elemento de su descripción en el diccionario. En este modelo es irrelevante.

El conjunto de símbolos no terminales incluye:

- el símbolo inicial *Logos*;
- el símbolo de un morfema de base *MB*;
- los símbolos del morfema de flexión *MFS* y *MFA*;
- los símbolos de vocales con señales de acento potencial:  $Va = \{\grave{a} \mid \grave{e} \mid \grave{i} \mid \grave{o} \mid \grave{u}\}$ . Las letras con acento grave no existen en el alfabeto español, y esto simplifica su comprensión como los no terminales, que actúan sólo en proceso de derivación de la cadena final. La codificación de las bases en símbolos de acento potencial es necesaria sólo para la clase 2 de sustantivos y las clases 3 y 4 de adjetivos. Ésta se realiza antes del trabajo del modelo;
- los símbolos de consonantes «fonéticos» *Z* y *Q*. La *Z* puede transformarse hacia *z* o *c* (*lápiz* vs. *lápices*), y *Q* hacia *c* o *qu* (*bloc* vs. *bloques*), dependiendo de sus contextos. La codificación de las últimas letras *z* y *c* es necesaria así antes de tiempo. Puede notarse que en el modelo formal de verbos de español la cantidad de consonantes «fonéticos» es mayor;
- los símbolos auxiliares:  $\{1 \mid 2 \mid 3 \mid \acute{\ } \mid \prime\}$  - para el número de una sílaba desde el fin de una palabra (los símbolos  $\acute{\ }$  y  $\prime$  marcan casos especiales de sílabas primera y segunda respectivamente); # - para el signo de principio y fin de la palabra.

El conjunto de símbolos terminales incluye:

- las vocales sin acento  $Vs = \{a \mid e \mid i \mid o \mid u \mid \ddot{u}\}$ ;
- las vocales con acento  $Vc = \{\acute{a} \mid \acute{e} \mid \acute{i} \mid \acute{o} \mid \acute{u}\}$ ;
- las consonantes *Co*, es decir, todas las letras restantes del alfabeto español. Por simplicidad, los digramas *ll* y *ch* se consideran como pares de consonantes independientes, este hecho no influye al trabajo de las reglas. En las reglas se usa también el signo  $Cr = Co \setminus s \setminus n$ , es decir todas las consonantes menos *s* y *n*.

**LAS REGLAS MORFOLÓGICAS FORMALES****Etapla 1. La conversión a una cadena de morfemas.** $(Po=sus) \vdash Logos \Rightarrow MT MFS \quad (1.sus)$  $(Po=adj) \vdash Logos \Rightarrow MT MFA \quad (1.adj)$ **Etapla 2. La concretización del morfo de la base.** $(L=l) \vdash MT \Rightarrow mt \quad (2.x)$ 

donde *l* y *mt* son significados concretos del lexema *L* y su base en el diccionario.

**Etapla 3s. Transición de la flexión del sustantivo a su morfo.** $(Cl=0) \vdash MFS \Rightarrow \# \quad (3s.0)$  $(Cl=1) \& (Num=sing) \vdash MFS \Rightarrow \# \quad (3s.1.sing)$  $(Cl=1) \& (Num=plur) \vdash MFS \Rightarrow s\# \quad (3s.1.plur)$  $(Cl=2) \& (Num=sing) \vdash MFS \Rightarrow 1\# \quad (3s.2.sing)$  $(Cl=2) \& (Num=plur) \vdash MFS \Rightarrow 2es\# \quad (3s.2.plur)$ **Etapla 3a. La transición de la flexión del adjetivo a su morfo.** $(Cl=0) \vdash MFA \Rightarrow \# \quad (3a.0)$  $(Cl=1) \& (Gen=masc) \& (Num=sing) \vdash MFA \Rightarrow o\# \quad (3a.1.masc.sing)$  $(Cl=1) \& (Gen=fem) \& (Num=sing) \vdash MFA \Rightarrow a\# \quad (3a.1.fem.sing)$  $(Cl=1) \& (Gen=masc) \& (Num=plur) \vdash MFA \Rightarrow os\# \quad (3a.1.masc.plur)$  $(Cl=1) \& (Gen=fem) \& (Num=plur) \vdash MFA \Rightarrow as\# \quad (3a.1.fem.plur)$  $(Cl=2) \& (Num=sing) \vdash MFA \Rightarrow \# \quad (3a.2.sing)$  $(Cl=2) \& (Num=plur) \vdash MFA \Rightarrow s\# \quad (3a.2.plur)$  $(Cl=3) \& (Num=sing) \vdash MFA \Rightarrow 1\# \quad (3a.3.sing)$  $(Cl=3) \& (Num=plur) \vdash MFA \Rightarrow 2es\# \quad (3a.3.plur)$  $(Cl=4) \& (Gen=masc) \& (Num=sing) \vdash MFA \Rightarrow 1\# \quad (3a.4.masc.sing)$  $(Cl=4) \& (Gen=fem) \& (Num=sing) \vdash MFA \Rightarrow 2a\# \quad (3a.4.fem.sing)$  $(Cl=4) \& (Gen=masc) \& (Num=plur) \vdash MFA \Rightarrow 2es\# \quad (3a.4.masc.plur)$  $(Cl=4) \& (Gen=fem) \& (Num=plur) \vdash MFA \Rightarrow 2as\# \quad (3a.4.fem.plur)$  $(Cl=5) \& (Gen=masc) \& (Num=sing) \vdash MFA \Rightarrow e\# \quad (3a.5.masc.sing)$  $(Cl=5) \& (Gen=fem) \& (Num=sing) \vdash MFA \Rightarrow a\# \quad (3a.5.fem.sing)$  $(Cl=5) \& (Gen=masc) \& (Num=plur) \vdash MFA \Rightarrow es\# \quad (3a.5.masc.plur)$  $(Cl=5) \& (Gen=fem) \& (Num=plur) \vdash MFA \Rightarrow as\# \quad (3a.5.fem.plur)$ **Etapla 4. La determinación del número de la sílaba.** $(n \mid s)l \Rightarrow 1(n \mid s) \quad (4.1)$  $(r \mid Z \mid Q)l \Rightarrow \gamma(Cr \mid Z \mid Q) \quad (4.2)$  $(Co \mid Z \mid Q)2 \Rightarrow 2(Co \mid Z \mid Q) \quad (4.3)$  $(Co \mid Z \mid Q)'' \Rightarrow \gamma(Co \mid Z \mid Q) \quad (4.4)$  $(Co \mid Z \mid Q)3 \Rightarrow 3(Co \mid Z \mid Q) \quad (4.5)$  $Vs \ 1 \Rightarrow 2 \ Vs \quad (4.6)$  $Vs \ ' \Rightarrow \gamma \ Vs \quad (4.7)$  $Vs \ 2 \Rightarrow 3 \ Vs \quad (4.8)$  $Vs \ '' \Rightarrow 3 \ Vs \quad (4.9)$  $Vs \ 3 \Rightarrow Vs \quad (4.10)$  $\#2 \Rightarrow \# \quad (4.11)$  $\#'' \Rightarrow \# \quad (4.12)$  $\#3 \Rightarrow \# \quad (4.13)$ **Etapla 5. La determinación del lugar del acento.** $\grave{a}1 \Rightarrow \acute{a} \quad (5.1.1)$  $\grave{e}1 \Rightarrow \acute{e} \quad (5.1.2)$  $\grave{i}1 \Rightarrow \acute{i} \quad (5.1.3)$  $\grave{o}1 \Rightarrow \acute{o} \quad (5.1.4)$  $\grave{u}1 \Rightarrow \acute{u} \quad (5.1.5)$  $\grave{a}2 \Rightarrow a \quad (5.2.1)$  $\grave{e}2 \Rightarrow e \quad (5.2.2)$  $\grave{i}2e \Rightarrow ie \quad (5.2.3)$  $\grave{i}2\bar{e} \Rightarrow i\bar{e} \quad (5.2.4)$  $\grave{o}2 \Rightarrow o \quad (5.2.5)$  $\grave{u}2e \Rightarrow ue \quad (5.2.6)$  $\grave{u}2\bar{e} \Rightarrow u\bar{e} \quad (5.2.7)$ 

( $\bar{e}$  es alguna letra diferente de la e)

 $\grave{a}'' \Rightarrow \acute{a} \quad (5.3.1)$  $\grave{e}'' \Rightarrow \acute{e} \quad (5.3.2)$  $\grave{i}'' \Rightarrow \acute{i} \quad (5.3.3)$  $\grave{o}'' \Rightarrow \acute{o} \quad (5.3.4)$  $\grave{u}'' \Rightarrow \acute{u} \quad (5.3.5)$  $\grave{a}3 \Rightarrow \acute{a} \quad (5.4.1)$  $\grave{e}3 \Rightarrow \acute{e} \quad (5.4.2)$  $\grave{i}3 \Rightarrow \acute{i} \quad (5.4.3)$  $\grave{o}3 \Rightarrow \acute{o} \quad (5.4.4)$  $\grave{u}3 \Rightarrow \acute{u} \quad (5.4.5)$ **Etapla 6. La apertura de consonantes fonéticas.** $Z\# \Rightarrow z\# \quad (6.1.1)$  $Ze \Rightarrow ce \quad (6.1.2)$  $Q\# \Rightarrow c\# \quad (6.2.1)$  $Qe \Rightarrow que \quad (6.2.2)$

**Etapa 7. El orden de aplicación de estas reglas es como sigue:**

Se buscan todas las posibilidades de la aplicación de las reglas de la etapa primera, en orden arbitrario. Cuando todas las posibilidades se acaban, se buscan las reglas de la etapa segunda, etc., hasta que todas las etapas se terminan.

**EJEMPLOS DE DERIVACIÓN DE SUSTANTIVOS**

Se darán unos ejemplos de derivación de formas de palabras para sustantivos. Cada línea de derivación se provee con el número de la regla correspondiente.

*(LUNES, (plur))*

Logos => MT MFS (1.sus)  
=> #lunes MFS (2.x)  
=> #lunes# (3s.0)  
=> lunes (7.1)

*(CASA, (plur))*

Logos => MT MFS (1.sus)  
=> #casa MFS (2.x)  
=> #casas# (3s.1.plur)  
=> casas (7.1)

*(VEZ, (plur))*

Logos => MT MFS (1.sus)  
=> #veZ MFS (2.x)  
=> #veZ2es# (3s.2.plur)  
=> #ve2Zes# (4.3)  
=> #v3eZes# (4.8)  
=> #3veZes# (4.5)  
=> #veZes# (4.13)  
=> #veces# (6.1.2)  
=> veces (7.1)

*(REGIÓN, (plur))*

Logos => MT MFS (1.sus)  
=> #regiòn MFS (2.x)  
=> #regiòn2es# (3s.2.plur)  
=> #regiò2nes# (4.3)  
=> #règiones# (5.2.5)  
=> regiones (7.1)

*(BLOC, (sing))*

Logos => MT MFS (1.sus)  
=> #bloQ MFS (2.x)  
=> #bloQ1# (3s.2.sing)  
=> #bloQ# (4.2)

=> #bl"oQ# (4.7)  
=> #b"loQ# (4.4)  
=> #"bloQ# (4.4)  
=> #bloQ# (4.12)  
=> #bloc# (6.2.1)  
=> bloc (7.1)

*(DÓLAR, (sing))*

Logos => MT MFS (1.sus)  
=> #dòlar MFS (2.x)  
=> #dòlar1# (3s.2.sing)  
=> #dòla'r# (4.2)  
=> #dòl"ar# (4.7)  
=> #dò"lar# (4.4)  
=> #dólar# (5.3.4)  
=> dólar (7.1)

*(DÓLAR, (plur))*

Logos => MT MFS (1.sus)  
=> #dòlar MFS (2.x)  
=> #dòlar2es# (3s.2.plur)  
=> #dòla2res# (4.3)  
=> #dòl3ares# (4.8)  
=> #dò3lares# (4.5)  
=> #dólares# (5.4.4)  
=> dólares (7.1)

*(JOVEN, (sing))*

Logos => MT MFS (1.sus)  
=> #jòven MFS (2.x)  
=> #jòven1# (3s.2.sing)  
=> #jòve1n# (4.1)  
=> #jòv2en# (4.6)  
=> #jò2ven# (4.3)  
=> #joven# (5.2.5)  
=> joven (7.1)

*(JOVEN, (plur))*

Logos => MT MFS (1.sus)  
=> #jòven MFS (2.x)  
=> #jòven2es# (3s.2.plur)  
=> #jòve2nes# (4.3)  
=> #jòv3enes# (4.8)  
=> #jò3venes# (4.5)  
=> #jóvenes# (5.4.4)  
=> jóvenes (7.1)

**EJEMPLOS DE DERIVACIÓN DE ADJETIVOS**

Los ejemplos de derivación de adjetivos españoles son más numerosos, en concordancia con la cantidad de sus clases morfológicas.

*(PEQUEÑO, (fem, plur))*

Logos => MT MFA	(1.adj)
=> #pequeñ MFA	(2.x)
=> #pequeñas#	(3a.1.fem.plur)
=> pequeñas	(7.1)

*(CIENTÍFICO, (masc, plur))*

Logos => MT MFA	(1.adj)
=> #científic MFA	(2.x)
=> #científicos#	(3a.1.masc.plur)
=> científicos	(7.1)

*(GRANDE, (fem, sing))*

Logos => MT MFA	(1.adj)
=> #grande MFA	(2.x)
=> #grande#	(3a.2.sing)
=> grande	(7.1)

*(COMÚN, (fem, plur))*

Logos => MT MFA	(1.adj)
=> #comùn MFA	(2.x)
=> #comùn2es#	(3a.3.plur)
=> #comù2nes#	(4.3)
=> #comunes#	(5.2.7)
=> comunes	(7.1)

*(BALADÍ, (masc, plur))*

Logos => MT MFA	(1.adj)
=> #baladì MFA	(2.x)
=> #baladì2es#	(3a.3.plur)
=> #baladíes#	(5.2.3)
=> baladíes	(7.1)

*(DÉBIL, (fem, sing))*

Logos => MT MFA	(1.adj)
=> #dèbil MFA	(2.x)
=> #dèbil1#	(3a.3.sing)
=> #dèbi1#	(4.2)
=> #dèb"il#	(4.7)
=> #dè"bil#	(4.4)
=> #débil#	(5.3.2)
=> débil	(7.1)

*(DIFÍCIL, (fem, plur))*

Logos => MT MFA	(1.adj)
=> #difícil MFA	(2.x)
=> #difícil2es#	(3a.3.plur)
=> #difíci2les#	(4.3)
=> #difíc3iles#	(4.5)
=> #difi3ciles#	(4.8)
=> #difíciles#	(5.4.3)
=> difíciles	(7.1)

*(FRANCÉS, (masc, sing))*

Logos => MT MFA	(1.adj)
=> #francès MFA	(2.x)
=> #francès1#	(3a.4.masc.sing)
=> #francè1s#	(4.1)
=> #francés#	(4.7)
=> francés	(7.1)

*(FRANCÉS, (fem, sing))*

Logos => MT MFA	(1.adj)
=> #francès MFA	(2.x)
=> #francès2a#	(3a.4.fem.sing)
=> #francè2sa#	(4.3)
=> #francesa#	(5.2.2)
=> francesa	(7.1)

*(ANDALUZ, (masc, sing))*

Logos => MT MFA	(1.adj)
=> #andaluZ MFA	(2.x)
=> #andaluZ1#	(3a.4.masc.sing)
=> #andaluZ#	(4.2)
=> #andal"uZ#	(4.7)
=> #anda"luZ#	(4.4)
=> #and3aluZ#	(4.9)
=> #an3daluZ#	(4.5)
=> #a3ndaluZ#	(4.5)
=> #andaluZ#	(4.10)
=> andaluz	(7.1)

*(GRANDOTE, (fem, plur))*

Logos => MT MFA	(1.adj)
=> #grandot MFA	(2.X)
=> #grandotas#	(3a.5.fem.plur)
=> grandotas	(7.1)

## VARIANTE CON LAS REGLAS INDEPENDIENTES DEL CONTEXTO

El modelo morfológico propuesto resultó inesperadamente complejo (64 reglas). Pero si no intentamos lograr la efectividad mayor de cómputo, el problema de síntesis al usar estas reglas es bastante simple: es suficiente probar, pasando todas las reglas en el ciclo anidado, por todas las etapas consecutivas, hasta que todas las reglas y las etapas se terminan. En el ejemplo de ANDALUZ claramente observamos cierta cantidad de operaciones superfluas (puramente tentativas).

El problema del análisis por estas reglas es más difícil, porque son dependientes del contexto y contienen transformaciones de los signos a cero.

Se puede uno preguntar, ¿existe un método para reducir las reglas dependientes del contexto hacia unas no dependientes? Sí, este método verdaderamente existe.

Los sustantivos en la clase 2 y los adjetivos en las clases 3 y 4 pueden tener dos variantes de la base, que se distinguen entre sí en una vocal acentuada o en la consonante final *z vs. c* o *c vs. qu*. Se guardarán en el diccionario separadamente ambos variantes *mt1* y *mt2* de la base como morfos concretos para cada lexema entre las clases indicadas. Estos variantes tienen sólo letras habituales del alfabeto español y son válidos cada uno en su subparadigma. Por ejemplo, para el sustantivo JOVEN los variantes son *joven* y *jóven*, para BLOC *bloc* y *bloqu*, para LÁPIZ *lápiz* y *lapic*, donde el primer morfo es para singular y el segundo para plural. Como ejemplos para adjetivos, COMÚN tiene los morfos *común* (para sing.) y *comun* (para plur.), e INGLÉS tiene *inglés* (para masc. sing.) e *ingles* (para todos los otros conjuntos de las características). En los casos de BALADÍ, DIFÍCIL, FÁCIL, etc. ambos morfos son iguales, pero esto no se contradice con el variante nuevo del modelo formal.

Entonces, el conjunto de reglas varía considerablemente:

La etapa 1 se mantiene.

La etapa 2 se complica:

$(L=1)\&(Po=sus)\&(Cl<2)$	1- MT => <i>mt</i>
$(L=1)\&(Po=sus)\&(Cl=2)\&(Num=sing)$	1- MT => <i>mt1</i>
$(L=1)\&(Po=sus)\&(Cl=2)\&(Num=plur)$	1- MT => <i>mt2</i>
$(L=1)\&(Po=adj)\&((Cl<3)\!(Cl=5))$	1- MT => <i>mt</i>
$(L=1)\&(Po=adj)\&(Cl=3)\&(Num=sing)$	1- MT => <i>mt1</i>
$(L=1)\&(Po=adj)\&(Cl=3)\&(Num=plur)$	1- MT => <i>mt2</i>
$(L=1)\&(Po=adj)\&(Cl=4)\&(Gen=masc)\&(Num=sing)$	1-MT => <i>mt1</i>
$(L=1)\&(Po=adj)\&(Cl=4)\&((Gen \neq masc)\!(Num \neq sing))$	1- MT => <i>mt2</i>

donde *l* es el signo de disyunción y  $\neq$  de desigualdad.

En las etapas 3s y 3a, en todas las reglas desaparecen los

signos no terminales #, 1 y 2.

Las etapas 4 y 5, las más complejas, desaparecen totalmente.

Por lo tanto, el modelo formal nuevo en vez de 64 reglas contiene sólo 32.

En lingüística general son conocidos los casos en los cuales un lexema tiene algunos morfos que la presentan en subparadigmas diferentes. El caso, en el cual un morfo de base no puede reducirse a otro por transformaciones simples y definidas, se llama el supletivismo de la base. En este caso el método de reducir ambos variantes al único se propone en el primer variante del modelo formal. Entonces el segundo variante puede llamarse el pseudo supletivismo.

El supletivismo permite en nuestro caso disminuir la cantidad de reglas en dos veces, pero el volumen de información guardada para los lexemas de las clases indicadas aumenta aproximadamente a dos veces.

En el nuevo variante del modelo, la cantidad de operaciones para generación de cualquier forma de palabra en todos los casos es igual a tres, como en los ejemplos:

*(DIFÍCIL, (fem, plur))*

*Logos => MT MFA*  
=> *dificil MFA*  
=> *dificiles*

*(ANDALUZ, (masc, sing))*

*Logos => MT MFA*  
=> *andaluz MFA*  
=> *andaluz*

En aplicaciones diferentes, hay casos en los cuales pueden ser más importantes las consideraciones de simplicidad de los algoritmos o del volumen de la memoria. Los dos variantes del modelo permiten seleccionar el variante adecuado. Para el análisis morfológico el variante con supletivismo es evidentemente preferible, porque las operaciones se reducen a una separación de una flexión como la parte última de la palabra y a una búsqueda de la parte restante entre los morfos de bases dentro del diccionario. Además en la separación y en la búsqueda se comparan cadenas de letras habituales porque las operaciones son técnicamente más fáciles.

## CONCLUSIÓN

Para los problemas de morfología de sustantivos y adjetivos del español, se dieron dos variantes equivalentes del conjunto de reglas rigurosas, que es necesario reproducir, en vía del análisis o de la síntesis morfológicos.

El modelo propuesto es un fragmento del modelo formal de la morfología del español en su totalidad, mientras que algunas reglas ya dadas son comunes para el modelo total.

## AGRADECIMIENTOS

Expreso mi agradecimiento a M. en C. Sofía N. Galicia Haro por su ayuda para corregir mi español y por otras notas valiosas, las cuales, espero, hayan hecho el texto de este artículo más legible.

## REFERENCIAS

- [1] Word para Windows 95. Guía del usuario. Microsoft Corp., 1995.
- [2] Power Translator Professional para Windows. Guía del usuario. Globalink, Inc., 1994.

[3] Mel'cuk, I. A. Morfología formal del verbo del español (en Ruso). En: Mashinnyj perevod i prikladnaja lingvistika, v. 10, 1974, Moscú.

[4] Bider I. G., I. A. Bolshakov, N. A. Eskova. Morfología formal del Ruso (en Ruso). Instituto de Lenguaje Ruso de Academia de Ciencias del URSS, Publ. #111, 112. Moscú, 1978.

[5] Bolshakov, I. A., A. A. Durnovo. Morfología formal de latín (en ruso). Instituto de Lenguaje Ruso de Academia de Ciencias del URSS, Publ. #124, 125. Moscú, 1979.

[6] Koskeniemi, K. Two-level model for morphological analysis. Proceedings of IJCAI, 1983, p. 683-685.

[7] Mel'cuk, Igor. Modelo de conjugación española. Voz y letra, 1993, v. 4, #1, p. 9-85

[8] Mel'cuk, I. A. Una experiencia en teorías de los modelos lingüísticos "Sentido  $\Leftrightarrow$  Texto" (en ruso). Publ. de "Nauka", Moscú, 1974.



*Igor A. Bolshakov. Nacionalidad rusa, obtuvo el doctorado en Ciencias Técnicas en el Instituto "Vypel". Sus áreas de interés son la lingüística aplicada, algoritmos y programación de computadoras especializadas en automatización de oficinas, en estandarización de software. Actualmente es profesor titular del Centro de Investigación en Computación del Instituto Politécnico Nacional, México.*



## DICE

PORTADA :  
Revista Iberoamerica de Computación

CONTRAPORTADA :  
Volumen. I Número. 1  
Volume. I. Number .1

006 Bruce H. Mc.Cormick,  
David A. Batte and Andrew T. Duchowski

014 José Ruiz. Shulcloper,  
Martín G. Chac y José F. Martínez  
Bases Conceptuales para una Teoría de Objetos Simbólicos

022 Gustavo Núñez and Matias Alvarado  
Contextual Knowledge Reasoning :  
Representation and Automatization

028 Igor A. Bolshakov  
El Modelo Morfológico Formal para la Síntesis de Sustantivos y Adjetivos en Español

038 J. Humberto Sossa and J. Luis Díaz de León  
Recognizing Planar Riding and Non-Riding Wire-Shapes

046 Sergio R. Sandoval  
Revisión de libros

048 Anuncios y Eventos  
Announcements and call for papers

PRIMERA DE FORROS  
Juan Manuel Ibarra Centro de Investigación  
en Computación-IPN México

PÁG. 1  
Volumen. I Número.1  
Volume. I. Number .1

006 Bruce H. Mc.Cormick,  
David A. Batte and Andrew T. Duchowski

014 José Ruiz. Shulcloper,  
Martín G. Chac y José F. Martínez  
Bases Conceptuales para una Teoría de Objetos Simbólicos

022 Gustavo Núñez and Matias Alvarado  
Contextual Knowledge Reasoning :  
Representation and Automatization

028 Igor A. Bolshakov  
El Modelo Morfológico Formal para la Síntesis de Sustantivos y Adjetivos en Español

038 J. Humberto Sossa and J. Luis Díaz de León  
Recognizing Planar Riding and Non-Riding Wire-Shapes

046 Sergio R. Sandoval  
Revisión de libros

048 Anuncios y Eventos  
Announcements and call for papers

PÁG. 2  
Línea 11  
invited several well know Mexican and foreign researchers in Computer Science and related fields,  
PÁG. 3  
Línea 9

Correo electrónico : Revista@pollux,cenac.ipn.mx.

PÁG. 26  
Matias Alvarado Mentado. Nacionalidad mexicana, es licenciado en Matemáticas por la Facultad de Ciencias de la Universidad Nacional Autónoma de México.

PÁG. 44  
Juan Luis Díaz de León Santiago. Nacionalidad mexicana, obtuvo el doctorado en el CINVESTAV del I.P.N, México con especialidad con ingeniería eléctrica. Profesor titular en el Centro de Investigación en el I.P.N.

PÁG. 52  
Línea 33  
[7]...

P ÁG. 53  
Línea 14 Los artículos deben ajustarse a las siguientes especificaciones :

## DEBE DECIR

Revista Iberoamericana de Computación

Volumen I Número 1  
Volume I Number 1

5 Bruce H. McCormick,  
David A. Batte and Andrew T. Duchowski

13 José Ruiz Shulcloper,  
Martín G. Chac y José F. Martínez  
Bases Conceptuales para una Teoría de Objetos Simbólicos

21 Gustavo Núñez and Matias Alvarado  
Contextual Knowledge and Belief :  
Representation and Reasoning

27 Igor A. Bolshakov  
El Modelo Morfológico Formal para Sustantivos y Adjetivos en el Español

37 J. Humberto Sossa and J. Luis Díaz de León  
Recognizing Planar Rigid and Non-Rigid Wire-Shapes

45 Sergio Sandoval Reyes  
Revisión de libros

47 Anuncios y Eventos  
Announcements and call for papers

Juan Manuel Ibarra Zannatha Centro de Investigación  
en Computación-IPN México

Volumen I Número 1  
Volume I Number 1

5 Bruce H. McCormick,  
David A. Batte and Andrew T. Duchowski

13 José Ruiz Shulcloper,  
Martín G. Chac y José F. Martínez  
Bases Conceptuales para una Teoría de Objetos Simbólicos

21 Gustavo Núñez and Matias Alvarado  
Contextual Knowledge and Belief :  
Representation and Reasoning

27 Igor A. Bolshakov  
El Modelo Morfológico Formal para Sustantivos y Adjetivos en el Español

37 J. Humberto Sossa and J. Luis Díaz de León  
Recognizing Planar Rigid and Non-Rigid Wire-Shapes

45 Sergio Sandoval Reyes  
Revisión de libros

47 Anuncios y Eventos  
Announcements and call for papers

invited several well known Mexican and foreign researchers in Computer Science and related fields,

Correo electrónico : Revista@pollux.cenac.ipn.mx.

Matias Alvarado Mentado. Nacionalidad mexicana, es licenciado en Matemáticas por la Facultad de Ciencias de la Universidad Nacional Autónoma de México.

PÁG. 44  
Juan Luis Díaz de León Santiago. Nacionalidad mexicana, obtuvo el doctorado en el CINVESTAV del I.P.N, México con especialidad en ingeniería eléctrica. Profesor titular en el Centro de Investigación en Computación en el I.P.N.

Línea 33  
[7] A. Rosenfeld. Digital Topology. *Amer.Math. Monthly*,  
*86* :621-630, 1979.  
[7] G. Matheron. *Random Sets and Integral Geometry*.  
John Wile and Sons, New York, 1995.  
[7] B. Vazquez, J.H. Sossa and J.L. Díaz de León  
Autoguided Vehicle Control using Expended Time  
B-splines. In *IEEE International Conference on Systems,  
Man and Cybernetics*, pages 2786 -2791, San Antonio  
Texas, October, 1994.  
[7] P. Perner, P. Wang and A. Rosefeld (Eds). *Advances  
in Structural and Syntactical Pattern Recognition*,  
Springer 1996.

Línea 14 Los artículos deben ajustarse a las siguientes especificaciones :