

Estudio Cuantitativo de la Sufijación Nominal en Español, en Textos Periodísticos

QUANTITATIVE STUDY OF NOMINAL SUFFIXATION IN SPANISH IN JOURNALISTIC TEXTS

Zulema Solana
GRUPO INFOSUR, UNR
Rosario, Argentina
zsolana@arnet.com.ar

Abstract

This article aims at presenting the possibilities of an automatic treatment to establish the frequency relative to the different noun-forming suffixes in journalistic texts and proposes to determine an index to measure suffixation in a text. We resort to the morphological analyzer Smorph, implemented as a labeler, to assign a category to all linguistic occurrences and then to Excel to extract the names from the list of categories rendered and to separate them into groups according to whether they carry a suffix assigning the noun category and to which suffix that is. Furthermore, frequencies relative to suffixes are established. This allows to determine an index, based on Information Theory, which is used in work on biodiversity.

Finally, the research proposes the same method to extend the comparison among standard text types in order to attain a general index.

Keywords: Suffixation index, Labeling.

Resumen

Este artículo se propone mostrar las posibilidades de un tratamiento automático para establecer la frecuencia relativa de los distintos sufijos formadores de nombres, en textos periodísticos y propone la determinación de un índice para medir la sufijación de un texto. Se recurre al analizador morfológico Smorph, implementado como etiquetador, para asignar categoría a todas las ocurrencias lingüísticas y luego a Excel para extraer los nombres de la lista de categorías logradas y para separarlos en conjuntos según lleven o no sufijo asignador de categoría nombre y según cuál sea éste. Además se establecen las frecuencias relativas de los sufijos. Todo esto permite proponer un índice, inspirado en la Teoría de la Información, al que se recurre en trabajos relativos a la biodiversidad.

La investigación, posteriormente, con el mismo método se propone ampliar la comparación de tipos de textos estándar para lograr un índice general

Palabras claves: índice de sufijación, etiquetado.

1. INTRODUCCION

Este artículo se propone mostrar las posibilidades de un tratamiento automático para establecer la frecuencia relativa de los distintos sufijos formadores de nombres, en textos periodísticos y lograr la determinación de un índice para medir la sufijación de un texto. Se recurre al analizador morfológico Smorph[1], implementado como etiquetador, para asignar categoría a todas las ocurrencias lingüísticas, y luego a Excel para extraer los nombres de la lista de categorías logradas y para separarlos en conjuntos según lleven o no sufijo asignador de categoría nombre y según cuál sea éste. Además, se consignan las frecuencias de los sufijos. Todo esto permite proponer un índice, inspirado en la Teoría de la Información, al que se recurre en trabajos relativos a la biodiversidad.

La investigación posteriormente, con el mismo método, se propone ampliar la comparación de tipos de textos estándar para lograr un índice general.

Se han realizado investigaciones sobre la distribución de nombres, adjetivos y verbos en textos científicos, técnicos y pedagógicos en español [2,3], pero no se han abordado las cuestiones como las que proponemos.

2. METODOLOGÍA

La metodología abarca los siguientes pasos:

- Selección de la muestra
- Etiquetado
- Pasaje a tabla Excel
- Filtrado automático de todos los nombres y de los nombres según sufijo.
- Volcado a tabla Excel del número total de aparición de cada sufijo.
- Cálculo, por medio de una función de Excel, de la frecuencia relativa de cada sufijo.
- Ordenamiento de la tabla, alfabéticamente y por frecuencia.

2.1. La muestra

Se selecciona una muestra de 8.000 palabras procedente de textos periodísticos que forman parte del banco de datos del grupo INFOSUR, es la trabajada por Beltrán [4].

2.2. Etiquetado

Mediante la declaración pertinente en cada uno de los archivos que componen el sistema SMORPH se implementa un etiquetador morfológico que hace uso del Módulo Post Smorph para resolver las ambigüedades mediante la definición de reglas específicas.

El software Smorph, analizador y generador morfosintáctico desarrollado en el Groupe de Recherche dans les Industries de la Langue (Universidad Blaise-Pascal, Clermont II) por Salah Ait-Mokhtar [3] realiza en una sola etapa la tokenización y el análisis morfológico. A partir de un texto de entrada se obtiene un texto lematizado con las formas correspondientes a cada lema (o a un subconjunto de lemas) con los valores correspondientes. Se trata de una herramienta declarativa, la información que utiliza está separada de la maquinaria algorítmica, en consecuencia, puede adaptarse a distintos usos. Con el mismo software se puede tratar cualquier lengua si se modifica la

información lingüística declarada en sus archivos.

Se ha trabajado con información pertinente para el francés [4], para el portugués [5] y para el español [6].

Smorph compila, minimiza y compacta la información lingüística que queda disponible en un archivo binario. Los códigos fuente se dividen en cinco archivos: Códigos ASCII, Rasgos, Terminaciones, Modelos y Entradas.

En el archivo **entradas**, se declaran los ítems léxicos acompañados por el modelo correspondiente. Este indicador de modelo oficia de enlace con el archivo **modelos**, en el que se especifica la información morfológica y las terminaciones que se requieren en cada ítem.

En el archivo **modelos**, se introduce la información correspondiente a los modelos de flexiones morfológicas. Se asocia a un conjunto de terminaciones el correspondiente conjunto de definiciones morfológicas. El esquema para definir los modelos es el siguiente:

```
<nombre_modelo> -<cantidad de caracteres a sustraer>
<terminación 1> <definición morfológica para terminación 1>
<terminación 2> <definición morfológica para terminación 2>
...
<terminación k> <definición morfológica para terminación k>
```

Se declara en primer lugar el nombre del modelo [7], luego la cantidad de caracteres que hay que sustraer a la forma lematizada. En tercer lugar se consigna la terminación, declarada previamente en el archivo **terminaciones**. La declaración morfológica corresponde a una cadena de caracteres sin espacios en blanco.

En el archivo **terminaciones** es necesario declarar todas las terminaciones que son necesarias para definir los modelos de flexión, se declaran una a continuación de otra, separadas por un punto.

Para construir los modelos se recurre a rasgos morfológico- sintácticos. En el archivo **rasgos**, se organizan jerárquicamente las etiquetas, por ejemplo, nombre, adjetivo, etc. En el archivo de códigos **ASCII** se especifican, entre otros, los caracteres separadores, las equivalencias entre mayúsculas y minúsculas. El archivo **data**, contiene los nombres de cada uno de los archivos descriptos anteriormente.

El módulo **post-smorph MPS** es un analizador que recibe en entrada una salida Smorph (en formato Prolog) y puede modificar las estructuras de datos recibidos. Ejecuta dos funciones principales: la **Recomposición** y la **Correspondencia**, que serán útiles para resolver las ambigüedades que resulten del análisis de Smorph.

En la muestra se obtuvieron 66 errores en etiquetado, esto es, 66 palabras a las que no se les asignó durante el proceso la etiqueta correcta. En términos relativos se traduce en un **99.17% de precisión global, con una confianza del 95% el intervalo correspondiente para el parámetro correspondiente a la precisión del sistema es (98.97%; 99.37%)**.

Análisis de Smorph [8]

```
'LA'.
[ 'el', 'EMS', 'det', 'TDET', 'art' ].
[ 'lo', 'EMS', 'cl', 'TPCL', 'nrfl' ].
'CRISIS'.
[ 'crisis', 'EMS', 'nom', 'GEN', 'fem', 'NUM', '___' ].
'EN'.
```

['en', 'EMS','prep'].
 'LA'.
 ['el', 'EMS','det', 'TDET','art'].
 ['lo', 'EMS','cl', 'TPCL','nrfl'].
 'PROVINCIA'.
 ['provincia', 'EMS','nom', 'GEN','fem', 'NUM','sg'].
 '.'.
 ['linsig', 'EMS','pun'].
 'En'.
 ['en', 'EMS','prep'].
 'un'.
 ['un', 'EMS','det', 'TINDF1','indfla'].
 'clima'.
 ['clima', 'EMS','nom', 'GEN','masc', 'NUM','sg'].
 'tenso'.
 ['tenso', 'EMS','adj', 'GEN','masc', 'NUM','sg'].
 ['tensar', 'EMS','v', 'MODOV','ind', 'PERS','1a', 'NUM','sg',
 'TPO','pres', 'TR','r', 'TC','c1'].
 ','.
 ['cc', 'EMS','coma'].

'Carlos'.
 ['Carlos', 'EMS','npr'].
 'Juárez'.
 ['Juárez', 'EMS','npr'].
 'se'.
 ['lo', 'EMS','cl', 'TPCRF','rflse'].
 'negó'.
 ['negar', 'EMS','v', 'MODOV','ind', 'PERS','3a', 'NUM','sg',
 'TPO','prets', 'TR','r', 'TC','c1'].
 'a'.
 ['a', 'EMS','prep'].
 'declarar'.
 ['declarar', 'EMS','v', 'MODOV','infin', 'TR','r', 'TC','c1'].
 '.'.
 ['linsig', 'EMS','pun'].

'El'.
 ['el', 'EMS','det', 'TDET','art'].

'ex'.
 ['ex', 'EMS','adv'].
 'gobernador'.
 ['gobernador', 'EMS','nom', 'GEN','masc', 'NUM','sg'].
 'santiagoño'.
 ['santiagoño', 'EMS','adj', 'GEN','masc', 'NUM','sg'].
 'debía'.
 ['deber', 'EMS','v', 'MODOV','ind', 'PERS','1a', 'NUM','sg',
 'TPO','imp', 'TR','r', 'TC','c2'].
 ['deber', 'EMS','v', 'MODOV','ind', 'PERS','3a', 'NUM','sg',
 'TPO','imp', 'TR','r', 'TC','c2'].
 'responder'.
 ['responder', 'EMS','v', 'MODOV','infin', 'TR','r', 'TC','c2'].
 'en'.

```
[ 'en', 'EMS', 'prep' ].
'los'.
[ 'el', 'EMS', 'det', 'TDET', 'art' ].
[ 'lo', 'EMS', 'cl', 'TPCL', 'nrfl' ].
'tribunales'.
[ 'tribunal', 'EMS', 'nom', 'GEN', 'masc', 'NUM', 'pl' ].
'locales'.
[ 'local', 'EMS', 'nom', 'GEN', 'masc', 'NUM', 'pl' ].
[ 'local', 'EMS', 'adj', 'GEN', '_', 'NUM', 'pl' ].
```

Análisis de Smorph/MPS (se aplican Reglas de Correspondencia)

```
'LA CRISIS EN'.
[ 'el crisis en', 'EMS', 'det_nom_prep' ].
```

```
'LA PROVINCIA'.
[ 'el provincia', 'EMS', 'det_nom' ].
```

```
'En un clima tenso'.
[ 'en un clima tenso', 'EMS', 'prep_det_nom_adj' ].
```

```
','.
[ 'cc', 'EMS', 'coma' ].
```

```
'Carlos'.
[ 'Carlos', 'EMS', 'npr' ].
```

```
'Juárez'.
[ 'Juárez', 'EMS', 'npr' ].
```

```
'se negó'.
[ 'lo negar', 'EMS', 'cl_v' ].
```

```
'a'.
[ 'a', 'EMS', 'prep' ].
```

```
'declarar'.
[ 'declarar', 'EMS', 'v', 'MODOV', 'infin' ].
```

```
'El'.
[ 'el', 'EMS', 'det', 'TDET', 'art' ].
```

```
'ex'.
[ 'ex', 'EMS', 'adv' ].
```

```
'gobernador'.
[ 'gobernador', 'EMS', 'nom' ].
```

```
'santiagoueño'.
[ 'santiagoueño', 'EMS', 'adj', 'GEN', 'masc', 'NUM', 'sg' ].
```

```
'debía'.
[ 'deber', 'EMS', 'v', 'MODOV', 'ind' ].
```

```
'responder'.
```

```
[ 'responder', 'EMS', 'v', 'MODOV', 'infin' ].
```

```
'en los tribunales locales'.
```

```
[ 'en el tribunal local', 'EMS', 'prep_det_nom_adj' ].
```

2.3. Pasaje a tabla Excel

Este archivo de texto se abre en Excel y se retienen las columnas que corresponden a la ocurrencia o palabra observada, el lema, la etiqueta y la EMS (etiqueta morfosintáctica). En los casos en que hay ambigüedad van a aparecer más de un lema y más de una etiqueta lo que lleva a construir nuevas columnas.

nro_secue	palab_obs	etiqu_1	etiqu_2	etiqu_correcta
1	LA	det	cl	det
2	CRISIS	nom	S/E	nom
3	EN	prep	S/E	prep
4	LA	det	cl	det
5	PROVINCIA	nom	S/E	nom
6	.	pun	S/E	pun
7	.	pun	S/E	pun
8	En	prep	S/E	prep
9	un	det	S/E	det
10	clima	nom	S/E	nom
11	tenso	adj	v	adj
12	,	coma	S/E	coma
13	Carlos	npr	S/E	npr
14	Juárez	npr	S/E	npr
15	se	cl	S/E	cl
16	negó	v	S/E	v
17	a	prep	S/E	prep
18	declarar	v	S/E	v
19	.	pun	S/E	pun
20	.	pun	S/E	pun
21	.	pun	S/E	pun
22	El	det	S/E	det
23	ex	adv	S/E	adv
24	gobernador	nom	S/E	nom
25	santiagoueño	adj	S/E	adj
26	debía	v	S/E	v
27	responder	v	S/E	v
28	en	prep	S/E	prep
29	los	det	cl	det
30	tribunales	nom	S/E	nom

Cuadro 1. Volcado de un fragmento del análisis de Smorph a una tabla Excel

2.4. Filtrado automático de los nombres

Recurriendo a la operación de filtrado se extrae de la lista total de palabras del texto a los nombres (cf. cuadro 2). Luego, en un proceso posterior, se los ordena alfabéticamente, aquí se muestra un fragmento del total de 1582 nombres y según el orden en que aparecen en el texto.

<u>palab_obs</u>	<u>etiqu_correcta</u>
CRISIS	nom
PROVINCIA	nom
clima	nom
gobernador	nom
tribunales	nom
responsabilidad	nom

Cuadro 2. Fragmento del inventario de nombres de la muestra

2.5. Filtrado de listas de nombres según sufijo

En una etapa posterior se declaran los sufijos de cada nombre, según consta en el fragmento siguiente.

clima	nom	base
gobernador	nom	dor
tribunales	nom	al
responsabilidad	nom	dad

Cuadro 3. Fragmento de la lista de nombres con asignación de sufijos

Existen dificultades para establecer el inventario de sufijos nominales ya se trate de hacer una segmentación adecuada de las palabras o de establecer si se trata de alomorfos o sufijos diferentes [cf. 7].

La primera dificultad fue marcar una línea divisoria entre palabras con y sin sufijo. Vamos a llamar “base” a los nombres a los que no se les asigna un sufijo: ya sea porque son nombres raíces como: “agua”, “año”, “crisis” o porque en su origen han tenido sufijo pero han perdido la relación con la palabra de la que derivan, como “muletilla”.

Para asignar los sufijos **-o** , **-a** y **-e** se estableció como condición, siguiendo a Lacuesta y Bustos (2000), que fueran de género inherente, ya que estos sufijos coinciden con otros flexionales que se adjuntan a nombres raíces, como perro/a, jefe/a. Se trató en todos los casos de nombres de acción salvo “intérprete” que es agente. Ejemplos: “ayuda”, “cosecha”, “culpa”, “acuerdo”, “apoyo”, “asalto”, “aumento”, “reproche”, “transporte”.

Los sufijos tienen en su mayoría género inherente masculino o femenino, hay sin embargo, algunos con una forma para el masculino y otra para el femenino como **-tor/a**, en ese caso consignamos sólo el masculino. En otros casos la forma masculina del sufijo (ej: terri-torio) no tiene vinculación con una femenina y a la inversa (convocatoria), por esa razón consignamos dos sufijos: **-torio** y **-toria**.

Los sufijos consignados son exclusivamente los que aparecen en la muestra y coinciden en gran parte con el inventario de Lacuesta y Bustos.

2.6. Volcar a tabla Excel el número total de aparición de cada sufijo

Se recurre nuevamente a la operación de filtrado para extraer a los nombres agrupados según el sufijo. Veamos un fragmento de los nombres con sufijo **-e** y con sufijo **-ción**.

informe	nom	e
intérprete	nom	e
parte	nom	e
reproche	nom	e
transporte	nom	e

Cuadro 4. Fragmento de la lista de nombres con sufijo -e

acción	nom	ción
aceptación	nom	ción
admiración	nom	ción
afección	nom	ción
afirmación	nom	ción
agrupación	nom	ción

Cuadro 5. Fragmento de la lista de nombres con sufijo -ción

2.7. Cálculo por medio de una función de la frecuencia relativa de cada sufijo

La frecuencia relativa de un sufijo es la proporción de veces que aparece dicho sufijo respecto del total de nombres que aparecen. Este cálculo se puede realizar mediante las operaciones matemáticas usuales en Excel.

2.8. Ordenar la tabla alfabéticamente y por frecuencia

Se presentará a continuación un cuadro con los sufijos aparecidos en la muestra, ordenados alfabéticamente y con la indicación de si tienen raíz verbal.

Total	1582	raíz verbal	raíz otra
a	50	50	0
ado	19	19	0
al	23	0	23
án	1	0	1
ancia	4	4	0
anga	1	1	0
ante	13	13	0
anza	1	1	0
ario	26	10	16
arra	1	0	1
ato	1	0	1
base	619	0	0
ción	185	181	4
dad	82	0	82
dor	27	20	7
e	16	16	0
eda	1	1	0
ena	2	0	2
encia	46	34	12
ente	1	0	1
ero	4	0	4

Total	1582	raíz verbal	raíz otra
eta	1	1	0
ete	1	1	0
eza	5	0	5
ia	4	0	4
ía	28	28	0
icia	1	1	0
icio	8	8	0
ida	4	4	0
ido	1	1	0
ija	1	1	0
il	5	0	5
ín	1	1	0
ina	2	0	2
io	32	23	9
ión	12	0	12
ismo	15	0	15
ista	12	8	4
ita	2	0	2
itis	1	1	0
ito	2	0	2
lleta	1	1	0
men	2	2	0
mento	3	0	3
miento	27	27	0
o	165	165	0
ón	1	0	1
or	9	9	0
oso	1	0	1
sión	29	25	4
ta	14	14	0
tad	3	0	3
te	3	3	0
tivo	11	0	11
to	7	4	3
tor	16	11	5
toria	5	4	1
torio	1	0	1
tud	2	2	0
uela	1	0	1
ura	24	7	17

Cuadro 6. Total de sufijos de la muestra ordenados alfabéticamente

Luego, como paso anterior a la aplicación de la fórmula del índice se calcula la frecuencia relativa de los sufijos.

3. OBSERVACIONES METODOLÓGICAS

Se hace necesario hacer las siguientes observaciones sobre el etiquetado y la asignación de sufijos.

El etiquetado guarda una estricta dependencia con los criterios usados para asignar categorías, a los que nos referiremos unas líneas más adelante. Esta asignación de categorías ya está realizada en nuestro diccionario lingüístico, es decir, en el archivo Entradas (de Smorph) que construimos en nuestra investigación, la que nos llevó a declarar 5512 lemas verbales, 9861 lemas nominales y

3059 lemas adjetivales, además de adverbios, preposiciones, pronombres, etc.

Como Smorph da como resultado una línea de análisis morfológico (una etiqueta) por cada análisis que se hace de cada ocurrencia lingüística, se producen ambigüedades de asignación categorial. En nuestro caso nos interesan particularmente las ambigüedades nombre/verbo y nombre/adjetivo.

En el caso de la ambigüedad nombre/ verbo, sólo se puede actuar después del análisis, ya que una ocurrencia lingüística, por ejemplo, *recuerdo* es analizada como nombre a partir de la declaración del lema nominal “recuerdo @n1 .” y como verbo a partir de la flexión del lema verbal “recordar”. Reglas de MPS permiten solucionar la ambigüedad.

En el caso de la ambigüedad nombre/adjetivo, siguiendo la propuesta de evitar una multiplicación innecesaria de ambigüedades que hace Hagégè [6], podemos tomar decisiones, como lo hemos hecho en el diccionario de Smorph (archivo “entradas”), antes de efectivizar el análisis automático. Por ejemplo, si aceptamos que un adjetivo puede ser núcleo del sintagma nominal:

-los tres argentinos

sólo figurará como adjetivo en las entradas y de ese modo será analizado por la herramienta informática que usamos.

4. ANÁLISIS DE RESULTADOS

El **39%** de las palabras **no** son derivadas, por lo tanto queda un **61%** de palabras con sufijos. Del total de sufijos, un **20%** corresponden a **-ción**, **16%** corresponden a **-o**, **8%** a **-dad**, **5%** a **-a**, **4%** a **-ncia** y **3,2%** a **-io** y **55** sufijos se reparten el **43,8** restante.

sufijos	%
ción	20%
o	16%
dad	8%
a	5%
ncia	4%
io	3%
restantes	44%

Cuadro 7. Distribución de los sufijos

Después de haber visto cómo se distribuyen cuantitativamente los sufijos se presentará un cuadro que pone en relación sufijo, tipo de género, tipo de base y significado general (se presenta a los de más uso).

sufijo	género inherente	base verbal	base adjetiva	base nominal	ejemplos
ción	+	+	-	-	aplicación, asignación, bendición, canción (nombres de acción)
o, a, e	+	+	-	-	ayuda, cosecha, culpa, acuerdo, apoyo, asalto, aumento, reproche, transporte.(en su mayoría nombres de acción)

sufijo	género inherente	base verbal	base adjetiva	base nominal	ejemplos
dad	+	-	+	-	responsabilidad, posibilidad, seguridad (nombres de cualidades)
ncia	+	+	+	-	ocurrencia, urgencia, ob servancia (nombres de acción)
io	+	+	-	-	equilibrio, disturbio (resultado de la acción verbal)
sión	+	+	-	-	conclusión, invasión, agresión (nombres de acción)
ía	+	-	+	+	burguesía , ciudadanía (nombres de condición o estatus)
dor/ dora	-	+	-	-	gobernador, seguidor, trabajador, vendedor (predominantemente agentes, instrumentos o locativos)
tor tor/ tora	+/-	+	-	-	director, inspector, sector, doctor (predominantemente agentes)
miento	+	+	-	-	tratamiento, emprendimiento, cumplimiento (nombres de acción)
ario ario/ aria	+/-	-	+	+	funcionario, secretario, escenario (agentes)
ura	+	+	+	-	altura, verdura (nombres de cualidades)
al	+	-	-	+	tribunal, hospital (significado locativo)
ado	+	+	-	-	estado, resultado, reinado (nombres de acción)
ismo	+	-	-	+	terrorismo, peronismo, partidismo ()

Cuadro 8, caracterización de los sufijos de mayor uso, ordenamiento según frecuencia salvo agrupamientos especiales

Tomando en cuenta los más usados, puede verse que en su mayoría el género masculino o femenino es inherente y que predomina la derivación nominal a partir de bases verbales.

La descripción hecha hasta el momento nos lleva a buscar un índice de la sufijación nominal teniendo en cuenta número de sufijos aparecidos en relación a la frecuencia de éstos. Es en la ecología y en la biología en general, donde encontramos el aprovechamiento del Índice de Shannon-Wiener, procedente de la Teoría de la Información para medir la biodiversidad. Lo adaptaremos para medir la sufijación.

Shannon-Wiener lo expresan del siguiente modo:

$$H = -\sum_{i=1}^S p_i \cdot \ln(p_i)$$

Siendo p_i la proporción de individuos de la especie i , es decir, $p=ni/N$ y s es la cantidad de especies distintas. En el caso que analizamos sería:

N =cantidad de nombres (1582)

n_i =cantidad de nombres con sufijo “Xi”

$p_i=n_i/N$

s =cantidad de sufijos distintos (61)

El valor del índice H para estos datos es 2,52.

Para comprender este valor es necesario considerar diferentes situaciones:

- Si por ejemplo, se tuviera en cada clase de sufijos la misma cantidad $(1582/61)=25$, este índice daría $H = 4,14$ (mayor que antes)
- Si de los 1582 nombres, 1520 fueran del primer sufijo y los restantes nombres 1 de cada 1, sería $H=0,327$ (menor que antes)
- Si en lugar de 61 sufijos se tuviesen 20 sufijos diferentes, igualmente distribuidos, $H=2,99$ (menor que el caso de todos los sufijos igualmente distribuidos)

Es decir, el índice H es mayor cuanto mayor es la cantidad de sufijos diferentes que aparece en el texto. Asimismo, para una misma cantidad de sufijos, será mayor el h en el caso en que todos tengan similares proporciones. Se tiene en cuenta no sólo la cantidad de sufijos distintos sino también cómo se distribuyen los nombres en esos tipos de sufijos.

Referencias

- [1] Aït-Mokhtar, Salah L'analyse présyntaxique en une seule étape. Tesis doctoral. Universidad Blaise-Pascal/GRIL, Clermont-Ferrand, 1998.
- [2] Soto, G., Martínez, R., Sadowsky, S. “ Verbos y sustantivos en textos científicos. Análisis de variación en un corpus de textos de ciencias aplicadas, naturales, sociales y humanidades” en PHILOLOGIA HISPALENSIS 19, 2005
- [3] Parodi, G., Venegas,R.. “Bucólico, aplicación computacional para el análisis de textos”, en Literatura y Lingüística n.15, Santiago 2004
- [4] Beltrán, C Modelización lingüística y análisis estadístico en el análisis automático de textos. Tesis de Doctorado, Facultad de Humanidades y Artes, 2008
- [5] Bès, G., Trouilleux, F. y Lamadon, L. “Verbal chunks extraction in French using limited resources” arXiv: cs.CL/0408060 v1
- [6] Hagège C. Analyse syntaxique automatique du portugais. Tesis de Doctorado GRIL, Univ. Blaise Pascal, 2000
- [7] Beltrán, C. “Comparación y evaluación de dos etiquetadores” Revista INFOSUR,N2, 2008
- [8] Lacuesta, R.S. y Bustos Guisbert, E. “La derivación nominal”, cap.69 de Bosque y Demonte (ed), Madrid 2000