

## ANÁLISIS AUTOMÁTICO MORFOLÓGICO CON LAS HERRAMIENTAS SMORPH Y NOOJ

### Automatic Morphological Analysis using SMORPH and NOOJ Tools

Celina Beltrán, Zulema Solana, Carolina Tramallino<sup>1</sup>  
Facultad de Humanidades y Artes UNR

#### Abstract

Morphology has had a great development within the scope of Computational Linguistics since it is the starting point to accomplish researches in other levels of representation. This time our focus of attention is on N (nouns) and A (adjectives) and the working logic of each tool while its common characteristics and peculiarities are indicated. The operations that can be performed using SMORPH are subtraction and concatenation at the end of a string, so, to formalize names we deal with lemmas and terminations. NooJ shares with Smorph the use of subtraction and concatenation at the end of a string, but besides that Nooj acts on different places within words –at the beginning or at the end- to be determined by the user who has the possibility to point out the direction, leftward or rightward. Nooj also can change or duplicate characters. To construct morphological grammars and NooJ dictionaries we resort to the already organized INFOSUR Dictionary (entries).

Key words: operation – subtraction – concatenation- grammars – dictionaries.

#### Resumen

La morfología ha tenido un gran desarrollo dentro de la Lingüística Computacional porque es la base para poder realizar investigaciones en los otros niveles de representación. En esta ocasión trabajaremos con N(nombres) y A (adjetivos), presentaremos la lógica operativa de cada herramienta marcando lo que tienen en común y sus particularidades. Las operaciones que efectúa SMORPH son la sustracción y la concatenación al final de una cadena, por

---

<sup>1</sup> Agradecemos en primer lugar al Dr Gabriel Bès, que fue quien inició la investigación en lingüística computacional en la UNR y en segundo término al Dr. Rodolfo Bonino, quien emprendió el estudio de NOOJ y nos auxilió generosamente.

lo tanto, para la formalización de la morfología de los nombres se trabajará con el lema y terminaciones. NooJ comparte con Smorph el uso de las operaciones de sustracción y concatenación al final de una cadena, pero además actúa sobre lugares a determinar por el usuario se trate del comienzo o fin de palabra y tiene la posibilidad de indicar la dirección (izquierda o derecha). No solamente sustrae o concatena sino además puede cambiar o duplicar caracteres. Para la construcción de las gramáticas morfológicas y los diccionarios de NOOJ recurriremos al diccionario (entradas) que ya ha trabajado el equipo INFOSUR

## **1. INTRODUCCIÓN**

La morfología ha tenido un gran desarrollo dentro de la Lingüística Computacional porque es la base para poder realizar investigaciones en los otros niveles de representación. En esta ocasión trabajaremos con N(nombres) y A (adjetivos), presentaremos la lógica operativa de cada herramienta marcando lo que tienen en común y sus particularidades.

## **2. Posibilidades del tratamiento morfológico inflexional del nombre en SMORPH**

### **2.1. La herramienta SMORPH**

#### **2.1.1. La herramienta en su conjunto**

Las operaciones que efectúa el programa son la sustracción y la concatenación al final de una cadena, por lo tanto, para la formalización de la morfología de los nombres se trabajará con el lema y terminaciones. La terminación es un concepto operacional puede coincidir con la desinencia, con la vocal temática más la desinencia o tener una estructura no comparable a la de la morfología tradicional.

El software Smorph es un analizador y generador morfosintáctico que fue desarrollado en el GRIL por Salah Aït-Mokhtar<sup>2</sup>, realiza en una sola etapa la tokenización y el análisis morfológico. Se trata de una herramienta declarativa porque la información que usa Smorph está separada de la maquinaria algorítmica, esto hace que se la pueda adaptar al uso que quiera darse, de modo tal que con el mismo software se puede tratar cualquier lengua siempre y cuando se modifique la información lingüística declarada en sus archivos.

El equipo de investigación al que pertenecemos, en una investigación ya concluida (INFOSUR) ha realizado la implantación en máquina mediante el software SMORPH de 4000 verbos del español, 6000 sustantivos, 3000 adjetivos, preposiciones, pronombres, etc.

---

<sup>2</sup> Bajo la dirección del Dr. Gabriel Bès.

Esta herramienta compila, minimiza y compacta la información lingüística de modo que quede disponible en un archivo binario. Los códigos fuente se dividen en cinco archivos:

- Códigos Ascii
- Rasgos
- Terminaciones
- Modelos
- Entradas

En el archivo “**entradas**” (que obra en este sistema como un “diccionario”), se ingresan los ítems léxicos acompañados por un indicador del modelo correspondiente. Este indicador de modelo hace de enlace con el archivo **modelos**, donde se especifica la información morfológica, género y número y las terminaciones que se requieren en cada ítem, es decir, el archivo “modelos” es la “gramática morfológica” del sistema. Ejemplo: Se define en Smorph el nombre *abuelo* y todas sus formas, *abuela*, *abuelos*, *abuelas*.

Archivo de **entradas**:

abuelo	@n4
--------	-----

Archivo de **modelos**:

@n4	-1	
	+o	masc/sg
	+a	fem/sg
	+os	masc/pl
	+as	fem/pl .

En el archivo **modelos**, se introduce la información correspondiente a los modelos de flexiones morfológicas. Un modelo de flexión agrupa todas las flexiones de una misma clase de palabras<sup>3</sup>. Esto se describe asociando 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<sup>4</sup>, en este ejemplo "@n4". Luego se declara la cantidad de caracteres que hay que extraer a la forma lematizada para obtener la raíz, en el ejemplo se debe extraer un solo caracter. Este valor debe ser una cifra entre 0 y 9 y estar precedida del signo "-". En tercer lugar se declara la terminación, la cual debe estar

<sup>3</sup> El modelo @n4 reúne a todos los nombres con la misma característica de *abuelo*, es decir, aquellos que al quitarle el último caracter y concatenarle cada una de las terminaciones definidas para dicho modelo dan origen a todas las formas: masculino/singular, masculino/plural, femenino/singular y femenino/plural.

<sup>4</sup> Los nombres de los modelos deben comenzar con el símbolo @.

declarada previamente en el archivo de **terminaciones**. La declaración morfológica corresponde a una cadena de caracteres sin espacios en blanco. En el ejemplo es *masc/sg* la información de género y número para la palabra abuelo (abuel +o). Es importante señalar que los modelos de flexión no necesariamente deben estar todos juntos en un mismo archivo.

Así, cuando el programa recoge la información de los archivos de este ejemplo, lee la palabra *abuelo*, le sustrae el último carácter y concatena cada una de las terminaciones expresadas en el modelo de flexión. de esta manera obtiene:

abuel + o = abuelo → (masc / sg)  
 abuel + os = abuelos → (masc / pl)  
 abuel + a = abuela → (fem / sg)  
 abuel + as = abuelas → (fem / pl)

#### Ej. Archivo **terminaciones**

<i>o .</i>
<i>a .</i>
<i>os .</i>
<i>as .</i>

En el archivo **terminaciones** es necesario declarar todas las terminaciones que son necesarias para definir los modelos de flexión. Si en la definición de un modelo se especifica una terminación no declarada en este archivo, el programa emite un mensaje de error. Las terminaciones se declaran una a continuación de otra, separadas por un punto. Es posible declarar una terminación vacía mediante el carácter "@" y una terminación distinguida asociando a una terminación la definición morfológica correspondiente.

Ejemplo de terminación distinguida para nombres femeninos singulares:  
*ción nom/fem/sg .*

Para construir los modelos se recurre a rasgos morfológico- sintácticos (categoría, género, número, etc). En el archivo **rasgos**, se organizan jerárquicamente las etiquetas, por ejemplo, nombre, adjetivo, etc. Asimismo, se puede incorporar la etiqueta que indica, por ejemplo, el tipo de nombre y se adicionan los rasgos de concordancia, género y número:

#### Ej: Archivo **rasgos**

EMS	etmorsin
nom	nombre
det	determinante
adj	adjetivo
TNOM	tiponombre
npr	nprop
nc	ncom .

GEN	género
	masc masc
	fem fem
NUM	número
	sg sing
	pl plur

A la derecha de cada etiqueta se escribe a modo de recordatorio alguna expresión que haga referencia a la abreviatura utilizada a la izquierda. Por ejemplo, *EMS* es la etiqueta morfosintáctica, en este ejemplo se definen tres: nombre (*nom*), adjetivo (*adj*) y determinante (*det*). Para los nombres además se define el tipo de nombre (*TNOM*), esto es, nombre propio (*npr*) y nombre común (*nc*). Luego, se definen el género (*GEN*) y el número (*NUM*).

En el archivo **códigos ASCII**, en el que 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.

Ej. salida de un texto analizado (solo la línea que corresponde al análisis solicitado)

'los'.

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

'hombres'.

[ 'hombre', 'EMS', 'nom', 'GEN', 'masc', 'NUM', 'pl' ].

'esperan'.

[ 'esperar', 'EMS', 'v', 'MODOV', 'ind', 'PERS', '3a', 'NUM', 'pl', 'TPO', 'pres', 'TR', 'r', 'TC', 'c1' ].

'en'.

[ 'en', 'EMS', 'prep' ].

'la'.

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

'plaza'.

[ 'plaza', 'EMS', 'nom', 'GEN', 'fem', 'NUM', 'sg' ].

## 2.1.2. Consideración pormenorizada de los modelos ( gramática morfológica)

### 2.1.2.1. Los sustantivos (N)

En 2.1.1. ejemplificamos con *abuelo*, caso en que se sustrae un caracter y se logra la raíz *abuel-*, a la que se agregan las terminaciones.

**a) Con esta lógica y sin recurrir a ningún otro procedimiento tenemos los modelos 1 a 11:**

@n1            -0  
           +@            nom/masc/sg  
           +s            nom/masc/pl .

(abrigo)

Como se trata de un nombre masculino cuyo plural se forma agregando –s sin ningún otro cambio, tenemos:

**ABRIGO-s**

**ABRIGO-s**

Con el mismo procedimiento tratamos a *mesa*:

@n2            -0  
           +@            nom/fem/sg  
           +s            nom/fem/pl .

**mesa**

**mesa-s**

Al igual que los nombres de forma común para el masculino y femenino.

@n3            -0  
           +@            nom/\_/sg  
           +s            nom/\_/pl  
                           **accionista**  
                           **accionista-s**

@n4            -1  
           +o            nom/masc/sg  
           +a            nom/fem/sg  
           +os            nom/masc/pl  
           +as            nom/fem/pl .

(perro)

En el modelo siguiente el plural se forma sobre el singular, pero como termina en consonante se agrega -es:

@n5            -0  
           +@            nom/masc/sg  
           +es            nom/masc/pl .

**acumulador**

**acumulador-es**

Con la misma estructura, pero femenino.

@n6            -0  
           +@            nom/fem/sg  
           +es            nom/fem/pl .

**pared**

**pared-es**

El siguiente modelo tiene la misma estructura que los dos anteriores, se lo distingue por poseer género indistinto:

@n7            -0

+@ nom/\_/sg  
 +es nom/\_/pl .

**bachiller**  
**bachiller-es**

El modelo n8 tiene terminaciones –a/-as para femenino:

@n8 -0  
 +@ nom/masc/sg  
 +es nom/masc/pl  
 +a nom/fem/sg  
 +as nom/fem/pl .

**acrededor**  
**acrededor-es**  
**acrededor-a**  
**acrededor-as**

El siguiente tiene género masculino y forma única para el número:

@n9 -0  
 +@ nom/masc/\_\_\_ .

**lunes**

El modelo 10 es igual al nueve pero los sustantivos que lo siguen son femeninos

@n10 -0  
 +@ nom/fem/\_\_\_ .

**crisis**

El once tiene forma única para los dos generos y los dos números:

@n11 -0  
 +@ nom/\_/\_

**aguafiestas**

### **b)problemas planteados por la presencia de tilde**

Desde el modelo 12 hasta el 25, 36 al 40, 43 al 47 la segmentación que hacemos no coincide con la que podría hacer la morfología tradicional (no computacional) porque **la presencia de la tilde en el singular** no permite tomarlo como base para el plural y en otros modelos para el femenino.

@n12 -2  
 +ón nom/masc/sg  
 +ones nom/masc/pl .

**sill-ón**  
**sill-ones**

@n13 -2  
 +ón nom/fem/sg  
 +ones nom/fem/pl .

**aberraci-ón**  
**aberraci-ones**

@n14 -2  
 +ón nom/masc/sg  
 +ones nom/masc/pl  
 +ona nom/fem/sg  
 +onas nom/fem/pl .

**anfitri-ón**

			<b>anfitri-ones</b> <b>anfitri-ona</b> <b>anfitri-onas</b>
@n15	-2	+ón +ones +onesa +onesas	nom/masc/sg nom/masc/pl nom/fem/sg nom/fem/pl .
			<b>bar-ón</b> <b>bar-ones</b> <b>bar-onesa</b> <b>bar-onesas</b>
@n16	-2	+án +anes	nom/masc/sg nom/masc/pl .
			<b>alacr-án</b> <b>alacr-anes</b>
@n17	-2	+én +enes	nom/masc/sg nom/masc/pl .
			<b>and-én</b> <b>and-enes</b>
@n18	-2	+ín +ines	nom/masc/sg nom/masc/pl .
			<b>plant-ín</b> <b>plant-ines</b>
@n19	-2	+ín +ines +ina +inas	nom/masc/sg nom/masc/pl nom/fem/sg nom/fem/pl .
			<b>bailar-ín</b> <b>bailar-ines</b> <b>bailarin-a</b> <b>bailarin-as</b>
@n20	-2	+ún +unes	nom/masc/sg nom/masc/pl .
			<b>bet-ún</b> <b>bet-unes</b>
@n21	-2	+ís +ises	nom/masc/sg nom/masc/pl .
			<b>an-ís</b> <b>an-ises</b>
@n22	-2		



	+ús	nom/masc/sg
	+uses	nom/masc/pl .
		<b>ob-ús</b>
		<b>ob-uses</b>
@n23	-4	
	+omen	nom/masc/sg
	+ómenes	nom/masc/pl .
		<b>abd-omen</b>
		<b>abd-ómenes</b>
@n24	-4	
	+umen	nom/masc/sg
	+úmenes	nom/masc/pl .
		<b>vol-umen</b>
		<b>vol-úmenes</b>
@n25	-6	
	+égimen	nom/masc/sg
	+egímenes	nom/masc/pl .
		<b>r-égimen</b>
		<b>r-egímenes</b>
@n36	-2	
	+és	nom/masc/sg
	+eses	nom/masc/pl
	+esa	nom/fem/sg
	+esas	nom/fem/pl .
		<b>marqu-és</b>
		<b>marqu-eses</b>
		<b>marqu-esa</b>
		<b>marqu-esas</b>
@n37	-4	
	+amen	nom/masc/sg
	+ámenes	nom/masc/pl .
		<b>ex-amen</b>
		<b>ex-ámenes</b>
@n38	-5	
	+argen	nom/masc/sg
	+árgenes	nom/masc/pl .
		<b>m-argen</b>
		<b>m-árgenes</b>
@n39	-4	
	+igen	nom/masc/sg
	+ígenes	nom/masc/pl .
		<b>or-igen</b>
		<b>or-ígenes</b>
@n40	-2	
	+ás	nom/masc/sg
	+ases	nom/masc/pl .
		<b>comp-ás</b>

		<b>comp-ases</b>
@n43	-4	
	+imen	nom/masc/sg
	+ímenes	nom/masc/pl .
		<b>cr-imen</b>
		<b>cr-ímenes</b>
@n44	-4	
	+ogan	nom/masc/sg
	+óganes	nom/masc/pl .
		<b>esl-ogan</b>
		<b>esl-óganes</b>
@n45	-5	
	+oquin	nom/masc/sg
	+óquines	nom/masc/pl .
		<b>esm-oquin</b>
		<b>esm-óquines</b>
@n46	-2	
	+és	nom/masc/sg
	+eses	nom/masc/pl .
		<b>Inter-és</b>
		<b>Inter-eses</b>
@n47	-4	
	+agen	nom/fem/sg
	+ágenes	nom/fem/pl .
		<b>Im-agen</b>
		<b>Im-ágenes</b>
<b><u>c) alternancia de Z/c</u></b>		
@n29	-1	
	+z	nom/masc/sg
	+ces	nom/masc/pl .
		<b>avestru-z</b>
		<b>avestru-ces</b>
@n30	-1	
	+z	nom/fem/sg
	+ces	nom/fem/pl .
		<b>cru-z</b>
		<b>cru-ces</b>
@n31	-1	
	+z	nom/_ /sg
	+ces	nom/_ /pl .
		<b>aprendi-z</b>
		<b>aprendi-ces</b>

**c) otros casos**

@n41	-1	
	+e	nom/masc/sg
	+es	nom/masc/pl
	+a	nom/fem/sg
	+as	nom/fem/pl .
		<b>nen-e</b>
		<b>nen-es</b>
		<b>nen-a</b>
		<b>nen-as</b>
@n26	-0	
	+@	nom/masc/sg .
		<b>ballet</b>
@n32	-0	
	+@	nom/masc/pl .
		<b>altibajos</b>
@n33	-0	
	+@	nom/fem/pl .
		<b>cosquillas</b>
@n35	-0	
	+@	nom/masc/sg
	+s	nom/masc/pl
	+esa	nom/fem/sg
	+esas	nom/fem/pl .
		<b>conde</b>
		<b>conde-s</b>
		<b>conde-sa</b>
		<b>conde-sas</b>

archivo terminaciones de N

@  
o  
a  
os  
as  
s  
es  
ón  
ones  
ona  
onas  
onesa  
onesas  
án  
anes  
ín

ines  
 ina  
 inas  
 ún  
 unes  
 én  
 enes  
 ís  
 ises  
 ús  
 uses  
 ás  
 ases  
 omen  
 ómenes  
 umen  
 imen  
 ímenes  
 argen  
 árgenes  
 igen  
 ígenes  
 ogan  
 óganes  
 z  
**ces**  
**e**  
**amen**  
 ámenes  
 égimen  
 egímenes  
 es  
 esas  
 eses  
 oquin  
 oquines  
 és  
 agen  
 ágenes

úmenes

### 3.Posibilidades del tratamiento morfológico inflexional del nombre en NOOJ<sup>5</sup>

#### 3.1.La herramienta NOOJ

##### 3.1.1.La herramienta en su conjunto

<sup>5</sup>cf.Rodolfo Bonino, Revista INFOSUR n6

Este programa de libre acceso fue creado en 2002, es una herramienta para el tratamiento de lenguas naturales y fue desarrollada por Max Silberstein. Actualmente cuenta con un foro de intercambio que puede consultarse en la página <http://groups.yahoo.com/group/nooj-info> en donde los usuarios pueden intercambiar opiniones y participar de congresos anuales que son organizados por su creador. El aporte personal reside en poder efectuar en un futuro, una adaptación de dicho programa que permita el análisis, la búsqueda de información y la comprobación de hipótesis lingüísticas en textos de español producidos por aprendientes de primera y de segunda lengua. Pero para llevar a cabo este desafío es fundamental la creación de gramáticas y diccionarios propios de la lengua española. Para tal fin recurrimos a la información lingüística declarada en los archivos de Smorph, el cual está diseñado para permitir que dicha información se modifique y adapte para el análisis de cualquier lengua. A partir de los modelos creados para sustantivos, adjetivos y verbos se procederá a la transformación de los mismos de acuerdo a los requerimientos algorítmicos de Nooj.

Nooj se puede descargar libremente del sitio web <http://www.nooj4nlp.net>. El usuario puede emplearlo, entre otras aplicaciones, para:

- Análisis de textos literarios.
- Investigación y extracción desde diarios o corpus técnicos.
- Formalización de fenómenos lingüísticos.
- Aplicaciones computacionales (análisis automático de textos).

Como especifica en su manual, incluye herramientas para crear y mantener fuentes lexicales así como gramáticas sintácticas y morfológicas. Estas maquinarias, al encontrarse integradas, permiten que se desarrollen las operaciones morfológicas sobre los ítems mientras se realiza una transformación sintáctica. Un ejemplo de ello puede ser realizar la transformación de oraciones en frases pasivas o controlar la concordancia sintáctica.

Puede procesar textos y corpus de cientos de archivos de texto, y lo más importante es que representa una herramienta multilingüe. Entre las 46 lenguas referidas encontramos: inglés, francés, portugués, alemán, italiano, ruso. En español no están cargados los datos lingüísticos correspondientes a gramáticas y diccionarios y por eso, actualmente, nuestro equipo de trabajo está ocupándose de crear los modelos y diccionarios correspondientes a las diferentes clases gramaticales, siguiendo las indicaciones de Max Silberstein, adaptando y actualizando la información lingüística declarada en los archivos con terminación txt de Smorph. Tal propósito es viable ya que las gramáticas y diccionarios de Nooj son fáciles de crear y de adaptar. Ambos programas incluyen herramientas que sirven para chequear, limpiar, adaptar, conservar y compartir los archivos.

➤ Tarea de reconocimiento:

Los usuarios pueden desarrollar extractores para identificar unidades semánticas en textos grandes, tales como nombres de personas, ubicaciones, fechas, expresiones técnicas o financieras, entre otras.

Los paradigmas Inflexión-derivación son formalizados como bibliotecas de gráficos (**grafos**) estructurados o reglas basadas en textos.

Las gramáticas sintácticas se utilizan para agregar anotaciones al texto o para desambiguar expresiones o estructuras.

Existen **dos tipos de recursos lingüísticos**

**a)Diccionarios** (archivos con terminación **.dic**). Asocian palabras o expresiones a:

- Un conjunto de información, tal como una categoría (ej. verbo),
- Uno o más paradigmas de inflexiones o derivaciones (por ejemplo cómo conjugar o nominalizar verbos).
- Una o más propiedades sintácticas (por ej: + transitivo).
- Una o más propiedades semánticas (por ejemplo clases distribucionales como “+Humano”)

Las propiedades léxicas pueden ser binarias o pueden estar expresadas como un atributo del par.

**b)Gramáticas:** se usan para representar una gran cantidad de fenómenos lingüísticos, desde niveles ortográficos y morfológicos a niveles sintagmáticos y sintácticos. Existen tres tipos:

- **Gramáticas Derivacionales o Inflexionales** (archivos con terminación **.nof** ) que se usan para representar las propiedades inflexionales, por ejemplo, conjugaciones o derivacionales como las nominalizaciones de las entradas léxicas. Estas entradas pueden realizarse de forma gráfica o mediante reglas.
- **Gramáticas lexicales, ortográficas, morfológicas o terminológicas** (archivos con terminación **.nom**). Se usan para representar conjuntos de tipos de palabras y asociarlas con información léxica. Por ejemplo: estandarizar la ortografía de palabras o de variantes, reconocer neologismos, asociar expresiones sinónimas.
- **Gramáticas semánticas o sintácticas.** (archivos con terminación **.nog**) Son utilizadas para reconocer o anotar expresiones en textos como ser: etiquetar tanto a frases nominales como a ciertas expresiones sintácticas o idiomáticas. Además se emplean para extraer expresiones o desambiguar palabras filtrando algunos ítems léxicos o anotaciones sintácticas en el texto.

El operador de disjunción permite comenzar varias búsquedas de extracción de información simultáneamente. En este caso son todas formas inflexionales de la misma palabra, pero también podrían localizarse variaciones ortográficas tales como: csar | czar | tsar | tza, nombres y variaciones tales como: New Cork City | Big apple | the city

Variaciones terminológicas: camcorder | video camera

Formas derivadas morfológicamente: Stalin | Stalinist | satlinism | destalinazation

O expresiones que representan los mismos conceptos: (credit | debit | ATM | visa) Card + Mastercard

### 3.1.1.1.Gramática inflexional de N en NOOJ

NooJ comparte con Smorph el uso de las operaciones de sustracción y concatenación al final de una cadena, pero además actúa sobre lugares a determinar por el usuario se trate del comienzo o fin de palabra y tiene la posibilidad de indicar la dirección (izquierda o derecha). No solamente sustrae o concatena sino además puede cambiar o duplicar caracteres.

Las operaciones básicas permiten crear operadores más específicos, que el autor del programa incorpora según los requerimientos de los usuarios; por ejemplo, cambiar una

vocal acentuada por una vocal inacentuada, en cuyo caso, el operador busca su argumento de izquierda a derecha.

Aclaración de los símbolos:

<E> significa que no ocurre ningún cambio.

<B> significa que se borra el último carácter.

<V> significa el cambio de z por c.

<A> significa que se quita el acento de la última vocal del lema.

<Á> significa que se agrega acento a la última vocal del lema.

<L5> significa que corremos el cursor cinco lugares.

“|” es el operador de disyunción

**-los comandos** se escriben en mayúscula y entre <>

**-las etiquetas** se escriben precedidas de /

-Si se **agregan números arábigos** después del comando, la operación se repite la cantidad de veces que indica el número. Por ejemplo, si ponemos <B3>, borra tres caracteres de izquierda a derecha; si ponemos <L2> mueve el cursor dos caracteres a la izquierda. Es posible combinar varios  
(cf. Bonino y Manual)

**ABRIGO** = <E>/masc+sg | s/masc+pl;

**MESA** = <E>/fem+sg | s/fem+pl;

**ACCIONISTA** = <E>/masc+sg | <E>/fem+sg | s/masc+pl | s/fem+pl;

**PERRO** = <E>/masc+sg | s/masc+pl | <B> a/fem+sg | <B> as/fem+pl;

**ACUMULADOR** = <E>/masc+sg | es/masc+pl;

**PARED** = <E>/fem+sg | es/fem+pl;

**BACHILLER** = <E>/masc+sg | <E>/fem+sg | es/masc+pl | es/fem+pl;

**ACREEDOR** = <E>/masc+sg | es/masc+pl | a/fem+sg | as/fem+pl;

**LUNES** = <E>/masc+sg | <E>/masc+pl;

**CRISIS** = <E>/fem+sg | <E>/fem+pl;

**SILLÓN** = <E>/masc+sg | <A> es/masc+pl;

**ABERRACIÓN** = <E>/fem+sg | <A> es/fem+pl;

Revista de Epistemología y Ciencias Humanas

**ANFITRIÓN** = <E>/masc+sg | <A> es/masc+pl | <A> a/fem+sg | <A> as/fem+pl;

**BARÓN** = <E>/masc+sg | <A> es/masc+pl | <A> esa/fem+sg | <A> esas/fem+pl;

**AVESTRUZ** = <E>/masc+sg | <V> es/masc+pl;

**CRUZ**=<E>/fem+sg | <V> es/fem+pl;

**APRENDIZ** = <E>/masc+sg | <E>/fem+sg | <V> es/masc+pl | <V> es/fem+pl;

**CONDE** = <E>/masc+sg | s/masc+pl | sa/fem+sg | sas/fem+pl;

**MARQUÉS** = <E>/masc+sg | <A> es/masc+pl | <A> a/fem+sg | <A> as/fem+pl;

**EXAMEN** = <E>/masc+sg | es <L5> <Á>/masc+pl;

**RÉGIMEN** = <E>/masc+sg | es <L7> <A> <R2> <Á>/masc+pl;

**COSQUILLAS** = <E>/fem+pl;

**AGUAFIESTAS** = <E>/masc+sg | <E>/fem+sg | <E>/masc+pl | <E>/fem+pl;

### 3.1.1.2.Diccionario de N en NOOJ

Primero va el lema (por ej. “ábaco”), luego una coma (,) y la categoría (N) después = (se flexiona como...) y “ABRIGO” es el modelo que sigue la flexión.

**Fragmento del diccionario de N**

**ábaco,N+FLX=ABRIGO**

**abanico,N+FLX=ABRIGO**

**abdomen,N+FLX=EXAMEN**

**abecedario,N+FLX=ABRIGO**

**abedul,N+FLX=ACUMULADOR**

**abeja,N+FLX=MESA**

**abejorro,N+FLX=ABRIGO**

**aberración,N+FLX=ABERRACIÓN**

**abertura,N+FLX=MESA**

**abeto,N+FLX=ABRIGO**

**abismo,N+FLX=ABRIGO**

**ablación,N+FLX=ABERRACIÓN**

**ablativo,N+FLX=ABRIGO**

**abnegación,N+FLX=ABERRACIÓN**

**abogado,N+FLX=PERRO**

**abogacía,N+FLX=MESA**

**abolladura,N+FLX=MESA**

**abono,N+FLX=ABRIGO**

**abordaje,N+FLX=ABRIGO**

**aborto,N+FLX=ABRIGO**



**abrazadera,N+FLX=MESA**  
**abrelatas,N+FLX=LUNES**  
**abreviatura,N+FLX=MESA**  
**abridor,N+FLX=ACUMULADOR**  
**abrigo,N+FLX=ABRIGO**  
**abril,N+FLX=ACUMULADOR**  
**abrochadora,N+FLX=MESA**  
**abrojo,N+FLX=ABRIGO**  
**absceso,N+FLX=ABRIGO**  
**abscisa,N+FLX=MESA**  
**ábside,N+FLX=ACCIONISTA**  
**absolutismo,N+FLX=ABRIGO**  
**absorción,N+FLX=ABERRACIÓN**  
**abstención,N+FLX=ABERRACIÓN**  
**abstinencia,N+FLX=MESA**  
**abuelo,N+FLX=PERRO**  
**abundancia,N+FLX=MESA**  
**aburrimiento,N+FLX=ABRIGO**  
**abuso,N+FLX=ABRIGO**  
**abyección,N+FLX=ABERRACIÓN**  
**acacia,N+FLX=MESA**  
**academia,N+FLX=MESA**  
**acantilado,N+FLX=ABRIGO**  
**ácaro,N+FLX=ABRIGO**  
**acceso,N+FLX=ABRIGO**  
**accidente,N+FLX=ABRIGO**  
**acción,N+FLX=ABERRACIÓN**  
**accionista,N+FLX=ACCIONISTA**  
**acecho,N+FLX=ABRIGO**  
**aceite,N+FLX=ABRIGO**  
**aceituna,N+FLX=MESA**  
**acelerador,N+FLX=ACUMULADOR**  
**acelga,N+FLX=MESA**  
**acento,N+FLX=ABRIGO**  
**acepción,N+FLX=ABERRACIÓN**  
**acequia,N+FLX=MESA**  
**aceptación,N+FLX=ABERRACIÓN**  
**acera,N+FLX=MESA**  
**acero,N+FLX=ABRIGO**  
**acertijo,N+FLX=ABRIGO**  
**acetato,N+FLX=ABRIGO**  
**acetileno,N+FLX=ABRIGO**  
**achaque,N+FLX=ABRIGO**  
**achicoria,N+FLX=MESA**  
**achira,N+FLX=MESA**  
**achura,N+FLX=MESA**  
**acicate,N+FLX=ABRIGO**

**ácido,N+FLX=ABRIGO**  
**acierto,N+FLX=ABRIGO**  
**acólito,N+FLX=ABRIGO**  
**acompañamiento,N+FLX=ABRIGO**  
**acontecimiento,N+FLX=ABRIGO**  
**acopio,N+FLX=ABRIGO**  
**acoplamiento,N+FLX=ABRIGO**  
**acorazado,N+FLX=ABRIGO**  
**acorde,N+FLX=ABRIGO**  
**acordeón,N+FLX=SILLÓN**  
**acotación,N+FLX=ABERRACIÓN**  
**acracia,N+FLX=MESA**  
**acreditación,N+FLX=ABERRACIÓN**  
**acreedor,N+FLX=ACREEDOR**  
**acritud,N+FLX=PARED**  
**acrobacia,N+FLX=MESA**  
**acróbata,N+FLX=ACCIONISTA**  
**acrópolis,N+FLX=CRISIS**  
**acta,N+FLX=MESA**  
**actitud,N+FLX=PARED**  
**actividad,N+FLX=PARED**  
**activo,N+FLX=ABRIGO**  
**acto,N+FLX=ABRIGO**  
**actor,N+FLX=ACUMULADOR**  
**actriz,N+FLX=CRUZ**  
**actualidad,N+FLX=PARED**  
**acuarela,N+FLX=MESA**  
**acuario,N+FLX=ABRIGO**  
**acueducto,N+FLX=ABRIGO**  
**acuerdo,N+FLX=ABRIGO**  
**acumulador,N+FLX=ACUMULADOR**  
**acusativo,N+FLX=ABRIGO**  
**acústica,N+FLX=MESA**  
**adagio,N+FLX=ABRIGO**  
**adalid,N+FLX=ACUMULADOR**  
**adaptación,N+FLX=ABERRACIÓN**  
**adefesio,N+FLX=ABRIGO**

### **3.2. Transformación de las entradas de nombres de SMORPH en diccionario de N de Nooj, de acuerdo a los requerimientos algorítmicos de Nooj.**

A modo de ejemplo se muestra un fragmento de la entrada de nombres en Smorph y el formato que debe tener para conformar el diccionario de nombres en Nooj (Figura 1).

Fragmento del diccionario de N	Fragmento de entradas de Smorph
<u>ábaco</u> , N+FLX=ABRIGO	<u>ábaco</u> @n1 .
<u>abanico</u> , N+FLX=ABRIGO	<u>abanico</u> @n1 .
<u>abdomen</u> , N+FLX=EXAMEN	<u>abdomen</u> @n23 .
<u>abecedario</u> , N+FLX=ABRIGO	<u>abecedario</u> @n1 .
<u>abedul</u> , N+FLX=ACUMULADOR	<u>abedul</u> @n5 .
<u>abeja</u> , N+FLX=MESA	<u>abeja</u> @n2 .

Figura 1: Fragmento del diccionario de N en Nooj y de las entradas de N en Smorph.

Es posible importar el archivo de texto conteniendo las entradas de Smorph en una planilla de cálculo Excel. De esta manera quedan las entradas en una columna y el modelo correspondiente en la columna siguiente como muestra parte del proceso en la figura 2 y la planilla de la figura 3.

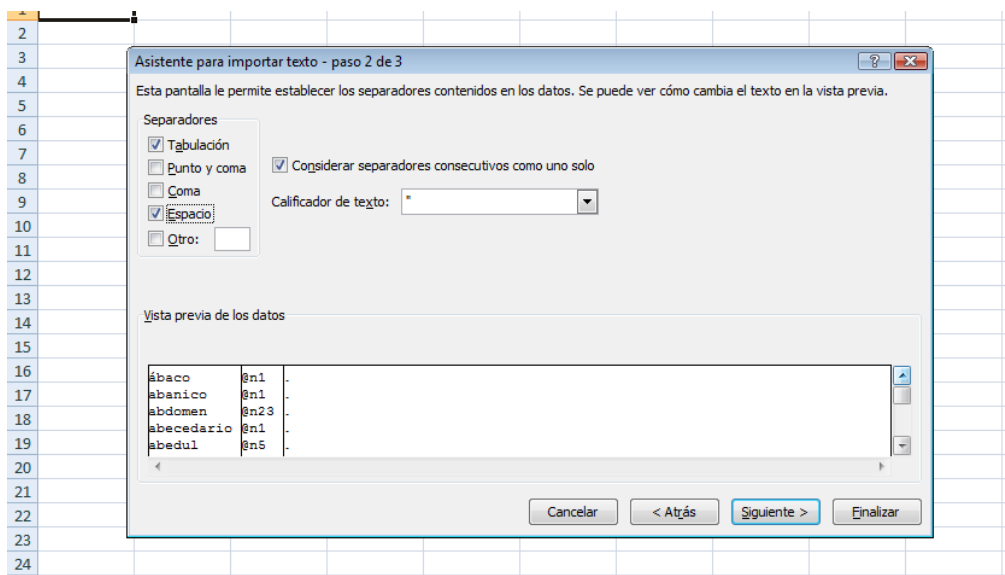


Figura 2: Importando los datos de entradas de Smorph en una planilla Excel.

	A	B	C	D
1	ábaco	@n1	.	
2	abanico	@n1	.	
3	abdomen	@n23	.	
4	abecedario	@n1	.	
5	abedul	@n5	.	
6	abeja	@n2	.	

Figura 3: Planilla Excel con las entradas de N en Smorph.

Teniendo la planilla de la figura 3, resta filtrar las celdas de cada modelo de Smorph y reemplazar por el modelo correspondiente de Nooj. Por ejemplo, filtrar las celdas del

modelo n1 y reemplazarlas por “ABRIGO”, luego filtrar las celdas n2 y reemplazar por “MESA” y así con cada modelo. Se van a realizar tantos filtros como modelos diferentes se tengan en Nooj.

Para completar el diccionario de Nooj resta agregar entre el lema y el nombre del modelo los caracteres “ ,N+FLX=”. Para realizar esto se inserta una columna entre la columna de lemas y la de modelos y se agrega en la primer fila la expresión “ ,N+FLX=” copiando automáticamente hacia abajo para completar toda la columna clickeando dos veces el vértice inferior derecho de la celda (Figura 4).

Por último, de manera de quitar los espacios y guardar nuevamente las entradas en un archivo de texto se utiliza la función concatenar de Excel para unir las 3 columnas en una sola y ésta llevarla a un archivo de texto (Figura 5). La columna final se guarda como archivo de texto (Figura 6).

	A	B	
1	ábaco	ABRIGO	
2	abanico	ABRIGO	
3	abdomen	EXAMEN	
4	abecedario	ABRIGO	
5	abedul	ACUMULADOR	
6	abeja	MESA	
7			

	A	B	C
1	ábaco	,N+FLX=	ABRIGO
2	abanico	,N+FLX=	ABRIGO
3	abdomen	,N+FLX=	EXAMEN
4	abecedario	,N+FLX=	ABRIGO
5	abedul	,N+FLX=	ACUMULADOR
6	abeja	,N+FLX=	MESA
7	abeiorro	,N+FLX=	ABRIGO

Figura 4: Planilla Excel con los cambios de modelos

CONCATENAR				
	A	B	C	D
1	ábaco	,N+FLX=	ABRIGO	+concatenar(A1;B1;C1)
2	abanico	,N+FLX=	ABRIGO	
3	abdomen	,N+FLX=	EXAMEN	
4	abecedario	,N+FLX=	ABRIGO	
5	abedul	,N+FLX=	ACUMULADOR	
6	abeja	,N+FLX=	MESA	

D1				
	A	B	C	D
1	ábaco	,N+FLX=	ABRIGO	ábaco,N+FLX=ABRIGO
2	abanico	,N+FLX=	ABRIGO	abanico,N+FLX=ABRIGO
3	abdomen	,N+FLX=	EXAMEN	abdomen,N+FLX=EXAMEN
4	abecedario	,N+FLX=	ABRIGO	abecedario,N+FLX=ABRIGO
5	abedul	,N+FLX=	ACUMULADOR	abedul,N+FLX=ACUMULADOR
6	abeja	,N+FLX=	MESA	abeja,N+FLX=MESA
7	abeiorro	,N+FLX=	ABRIGO	abeiorro,N+FLX=ABRIGO

Figura 5: Concatenación de las columnas para obtener el diccionario de N en Nooj.

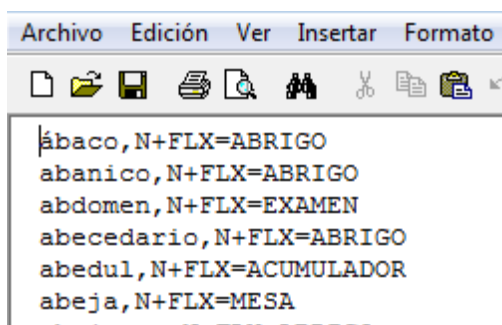


Figura 6: Diccionario final de N en Nooj.

#### 4. Posibilidades del tratamiento morfológico inflexional del adjetivo en SMORPH

@a1	-0	
+@		adj/masc/sg
+s		adj/masc/pl .
		<b>peristáltico</b>
		<b>peristáltico-s</b>
@a2 (ex3)	-0	
+@		adj/_/sg
+s		adj/_/pl .
		<b>recurrente</b>
		<b>Recurrente-s</b>
@a3(ex4)	-1	
+o		adj/masc/sg
+a		adj/fem/sg
+os		adj/masc/pl
+as		adj/fem/pl .

**flac-o****flac-a****flac-os****flac-as**

@a4 (ex7) -0

+@ adj/\_/sg

+es adj/\_/pl .

**ágil****ágil-es**

@a5 (ex8) -0

+@ adj/masc/sg

+es adj/masc/pl

+a adj/fem/sg

+as adj/fem/pl .

**opositor****opositor-es****opositor-a****opositor-as**

@a6 (ex 9) -2

+ín adj/masc/sg

+ines adj/masc/pl

+ina adj/fem/sg

+inas adj/fem/pl .

**cantar-ín****cantar-ines**

		<b>cantar-ina</b>
		<b>cantar-inas</b>
@a7 (ex14)	-2	
+ón		adj/masc/sg
+ones		adj/masc/pl
+ona		adj/fem/sg
+onas		adj/fem/pl .
		<b>santurr-ón</b>
		<b>santurr-ones</b>
		<b>santurr-ona</b>
		<b>santurr-onas</b>
@a8 (ex36)	-2	
+és		adj/masc/sg
+eses		adj/masc/pl
+esa		adj/fem/sg
+esas		adj/fem/pl .
		<b>aragon-és</b>
		<b>aragon-eses</b>
		<b>aragon-esa</b>
		<b>aragon-esas</b>
@a9 (ex50)	-1	
+y		adj/_/sg
+ies		adj/_/pl .
		<b>hipp-y</b>
		<b>hipp-ies</b>

@a10(ex51) -2

+án adj/masc/sg

+ana adj/fem/sg

+anes adj/masc/pl

+anas adj/fem/pl .

**harag-án**

**harag-anes**

**harag-ana**

**harag-anas**

@a11 (ex52) -1

+z adj/\_/sg

+ces adj/\_/pl

**feli-z**

**feli-ces**

@a12 (ex53) -2

+ón adj/\_/sg

+ones adj/\_/pl .

**marr-ón**

**mar-ones**

@a13 (ex54) -2

+ún adj/\_/sg

+unes adj/\_/pl .

**com-ún**

**com-unes**



## 5. Posibilidades del tratamiento morfológico inflexional del adjetivo en NOOJ

### 5.1. Gramática inflexional de A en NOOJ

**PERISTÁLTICO** = <E>/masc+sg | s/masc+pl;

**FLACO** = <E>/masc+sg | s/masc+pl | <B> a/fem+sg | <B> as/fem+pl;

**RECURRENTE** = <E>/masc+sg | <E>/fem+sg | s/masc+pl | s/fem+pl;

**ÁGIL** = <E>/masc+sg | <E>/fem+sg | es/masc+pl | es/fem+pl;

**OPOSITOR** = <E>/masc+sg | es/masc+pl | a/fem+sg | as/fem+pl;

**HARAGÁN** = <E>/masc+sg | <A> es/masc+pl | <A> a/fem+sg | <A> as/fem+pl;

**ANDALUZ** = <E>/masc+sg | <E> a/fem+sg | <V> es/masc+pl | <E> as/fem+pl;

**COMÚN** = <E>/masc+sg | <A> es/masc+pl | <E>/fem+sg || <A> es/fem+pl;

### 5.2. Diccionario de A en NOOJ

**abatido, AJ+FLX=FLACO**

**aberrante, AJ+FLX=RECURRENTE**

**abierto, AJ+FLX=FLACO**

**abigarrado, AJ+FLX=FLACO**

**abismal, AJ+FLX=ÁGIL**

**abominable, AJ+FLX=RECURRENTE**

**abortivo, AJ+FLX=FLACO**

**abovedado, AJ+FLX=FLACO**

**abrasivo, AJ+FLX=FLACO**

**abriboca, AJ+FLX=RECURRENTE**

**abrupto, AJ+FLX=FLACO**

**absoluto, AJ+FLX=FLACO**

**absorbente, AJ+FLX=RECURRENTE**

**absorto, AJ+FLX=FLACO**

**abstemio, AJ+FLX=FLACO**

**abstracto, AJ+FLX=FLACO**

**abstraído, AJ+FLX=FLACO**

**abstruso, AJ+FLX=FLACO**

**abusivo, AJ+FLX=FLACO**

**abyecto, AJ+FLX=FLACO**

**académico, AJ+FLX=FLACO**

**acalorado, AJ+FLX=FLACO**

**acanalado, AJ+FLX=FLACO**

**accesorio, AJ+FLX=FLACO**

**accidental, AJ+FLX=ÁGIL**

**acéfalo,AJ+FLX=FLACO**  
**aceitoso,AJ+FLX=FLACO**  
**aceptable,AJ+FLX=RECURRENTE**  
**acérrimo,AJ+FLX=FLACO**  
**achacoso,AJ+FLX=FLACO**  
**acaparrado,AJ+FLX=FLACO**  
**achinado,AJ+FLX=FLACO**  
**achiquillado,AJ+FLX=FLACO**  
**aciago,AJ+FLX=FLACO**  
**ácido,AJ+FLX=FLACO**  
**aconsejable,AJ+FLX=RECURRENTE**  
**ácrata,AJ+FLX=RECURRENTE**  
**acrisolado,AJ+FLX=FLACO**  
**acromático,AJ+FLX=FLACO**  
**activo,AJ+FLX=FLACO**  
**activista,AJ+FLX=RECURRENTE**  
**actual,AJ+FLX=ÁGIL**  
**acuático,AJ+FLX=FLACO**  
**acuífero,AJ+FLX=FLACO**  
**acuoso,AJ+FLX=FLACO**  
**acústico,AJ+FLX=FLACO**  
**adepto,AJ+FLX=FLACO**  
**adhesivo,AJ+FLX=FLACO**  
**adicional,AJ+FLX=ÁGIL**  
**adicto,AJ+FLX=FLACO**  
**adinerado,AJ+FLX=FLACO**  
**adiposo,AJ+FLX=FLACO**  
**aditivo,AJ+FLX=FLACO**  
**adjunto,AJ+FLX=FLACO**  
**adoptivo,AJ+FLX=FLACO**  
**adúltero,AJ+FLX=FLACO**  
**adulto,AJ+FLX=FLACO**  
**adusto,AJ+FLX=FLACO**  
**advenedizo,AJ+FLX=FLACO**  
**adversario,AJ+FLX=FLACO**  
**adverso,AJ+FLX=FLACO**  
**adyacente,AJ+FLX=RECURRENTE**  
**aerodinámico,AJ+FLX=FLACO**  
**aeroespacial,AJ+FLX=ÁGIL**  
**aeronaval,AJ+FLX=ÁGIL**  
**afable,AJ+FLX=RECURRENTE**  
**afectuoso,AJ+FLX=FLACO**  
**aferente,AJ+FLX=RECURRENTE**  
**afestonado,AJ+FLX=FLACO**  
**afiligranado,AJ+FLX=FLACO**  
**afortunado,AJ+FLX=FLACO**  
**africado,AJ+FLX=FLACO**

**afuereño,AJ+FLX=FLACO**  
**agalludo,AJ+FLX=FLACO**  
**agente,AJ+FLX=RECURRENTE**  
**ágil,AJ+FLX=ÁGIL**  
**aglomerante,AJ+FLX=RECURRENTE**  
**agnóstico,AJ+FLX=FLACO**  
**agrario,AJ+FLX=FLACO**  
**agresivo,AJ+FLX=FLACO**  
**agreste,AJ+FLX=RECURRENTE**  
**agridulce,AJ+FLX=RECURRENTE**  
**agrio,AJ+FLX=FLACO**  
**agropecuario,AJ+FLX=FLACO**  
**aguado,AJ+FLX=FLACO**  
**agudo,AJ+FLX=FLACO**  
**aguileño,AJ+FLX=FLACO**  
**ahijado,AJ+FLX=FLACO**  
**ahuesado,AJ+FLX=FLACO**  
**airoso,AJ+FLX=FLACO**  
**ajeno,AJ+FLX=FLACO**  
**alado,AJ+FLX=FLACO**  
**albino,AJ+FLX=FLACO**  
**alcachofero,AJ+FLX=FLACO**  
**alcalino,AJ+FLX=FLACO**  
**alcista,AJ+FLX=RECURRENTE**  
**aldeano,AJ+FLX=FLACO**

## BIBLIOGRAFÍA

Aït Mokthar, Salah 1998 *L'analyse présyntaxique en une seule etape* tesis doctoral dirigida por Gabriel G. Bès en el GRIL Université Blaise-Pascal, Francia

Aït Mokthar, Salah 1995 *SMORPH : Guide d'utilisation. Rapport technique*, Universidad Blaise Pascal/GRIL, Clermont-Fd.

Aït-Mokhtar, Salah y Rodrigo Mateos, José Lázaro. 1995 Segmentación y análisis morfológico de textos en español utilizando el sistema SMORPH. *SEPLN*, 17, 29-41.

Silberztein Max, 2003-. NooJ Manual. Available for download at:  
[www.nooj4nlp.net](http://www.nooj4nlp.net)