

## **El Sintagma adjetivo núcleo**

### **The Nucleus Adjective Syntagma**

**Bárbara Méndez, Mariana Morosano, Andrea Ortiz**

Universidad Nacional de Rosario

Facultad de Humanidades y Artes

Rosario, Argentina

barbaramendez555@hotmail.com, maia\_47@ubbi.com, andrea\_ortiz24@hotmail.com

### **Resumen**

En este trabajo realizaremos la descripción, formalización y la implantación en máquina del sintagma adjetivo núcleo (sadjn). Para la implantación, nos valdremos de Smorph y MPS, dos herramientas informáticas que se utilizan para el procesamiento de textos. Smorph realiza el análisis morfológico y MPS trabaja sobre gramáticas locales. Ambas integran un módulo, de modo que la salida de Smorph es la entrada de MPS.

La descripción y formalización del sadjn se enmarcan en el Paradigma 5P desarrollado en el GRIL (Groupe de Recherche dans les Industries de la Langue) Universidad Blaise-Pascal de Clermont-Ferrand (Francia) bajo la dirección del Dr. Gabriel Bès.

**Palabras claves:** Descripción y formalización, Paradigma 5P, implantación en máquina.

### **Abstract**

In this work we will present the description, formalization and the implantation in machine of nucleus adjective syntagma (sadjn). For the implantation, we will use Smorph and MPS, two computer science tools used for word processing. Smorph operates the morphological analysis and MPS works on local grammars. Both integrate a module, so that the output of Smorph is the input to MPS.

The description and formalization of sadjn are framed within the 5P Paradigm developed by GRIL (Groupe de Recherche dans les Industries de la Langue), University Blaise-Pascal at Clermont-Ferrand (France), under the direction of Dr Gabriel Bès.

**Key words:** Description, formalization, implantation in machine.

## 1. Introducción

Este trabajo se propone la descripción, formalización y la implantación en máquina del sintagma adjetivo núcleo (sadjn). Utilizaremos Smorph y MPS, dos herramientas informáticas que sirven para el procesamiento de textos. Smorph realiza el análisis morfológico y MPS trabaja sobre gramáticas locales. Ambas integran un módulo, de modo que la salida de Smorph es la entrada de MPS.

Realizaremos la descripción y formalización del sadjn en el marco del Paradigma 5P, desarrollado en el GRIL (Groupe de Recherche dans les Industries de la Langue) Universidad Blaise-Pascal de Clermont-Ferrand (Francia) bajo la dirección del Dr. Gabriel Bès.<sup>4</sup> y continuado en la Universidad Nacional de Rosario en Seminarios dictados durante 2004 y 2005.

Es importante destacar que este trabajo integra el *Proyecto Infosur: investigación y desarrollo* dirigido por la Dra. Zulema Solana, con la asesoría científica del Dr. Gabriel G. Bès.

El Proyecto Infosur apunta a la implantación en máquina de:

- una base de datos morfológicos del español
- una base de datos sintácticos del español y
- una base de datos léxicos del español.

Se espera conseguir un prototipo informático que posibilite el tratamiento automático, morfológico y sintáctico de textos del español.

Con respecto a los sintagmas núcleos, considerando las categorías morfosintácticas que los integran, podemos determinar dónde comienzan, dónde terminan, cómo están compuestos y cuál es su núcleo. El orden en que se presentan las categorías y la posibilidad de combinación están restringidos.

Un sintagma núcleo es un bloque casi inseparable; en ejemplos como (1) *los escuché*; (2) *los verdes prados*; (3) *muy atento* se tiene respectivamente un sintagma verbal núcleo (svn), un sintagma nominal núcleo (snn) y un sintagma adjetivo núcleo (sadjn).

Los sintagmas núcleos son los *chunks* que propone Abney [1], dando razones prosódicas y psicolingüísticas; considera que permiten un análisis automático del texto con menores dificultades. Ait-Mokhtar y Chanod [2] y continuadores han hecho investigaciones de análisis sintáctico automático recurriendo a la plataforma XIP que analiza en *chunks* o sintagmas núcleos [3].

El análisis en sintagmas núcleos, justificado en el ámbito del Paradigma 5P, hace posible una reducción significativa de la ambigüedad de la categorización morfo-sintáctica por medio de la concatenación de las expresiones internas. También resulta un paso en la tokenización que es un avance para lograr la representación semántica. El análisis que realiza 5P focaliza la estructura interna de los chunks [4], [5] y [6].

## 2. ANÁLISIS LINGÜÍSTICO DE LOS SADJN

En este apartado desarrollaremos características generales del sadjn, es decir estudiaremos algunos puntos que consideramos fundamentales como: qué entendemos por sadjn, cuál es su núcleo, qué categorías lo integran y qué combinaciones admite. Es importante destacar que estos puntos del sadjn son pertinentes para la clasificación de los adjetivos.

En primer lugar entendemos al sadjn como el segmento que empieza con el primer elemento del sintagma adjetivo y finaliza en su núcleo. Así en ocurrencias como *muy hermoso*, el sadjn comienza

---

<sup>4</sup> Gabriel G. Bès es el asesor científico desde 2004 de nuestra investigación en Lingüística Computacional INFOSUR.

en el adverbio y termina en el núcleo. En el caso de *hermoso* el *sadjn* está conformado sólo por el núcleo.

El *sadjn* lleva siempre como núcleo un adjetivo, a diferencia del *snn* (sintagma nominal núcleo) que puede llevar como núcleo una categoría distinta a *n* (nombre), así en *Los tres llegaron*, el núcleo del *snn* es el numeral.

Los adjetivos pueden aparecer solos, sin que los anteceda ningún elemento.

Las categorías que integran el *sadjn* son: adjetivos, adverbios e indefinidos.

## 2.1. Las categorías

Se trabajará con las siguientes categorías:

adj

adj1

adj2

adv

indf

adj. 1: son los que pueden ir solos o antecidos de adverbios e indefinidos.

Por ejemplo:

*hermoso* (adj1)

*complicado* (adj1)

*muy hermoso* (adv +adj1)

*muy complicado* (adv + adj1)

*demasiado hermosa* (indf+ adj1)

*demasiado complicado* (indf +adj1)

adj 2: deben ir solos

Por ejemplo:

***moral* (adj 2)**

***universitario* (adj 2)**

*internacional.* (adj.2)

adv: *muy, tan, etc.*

indf: *demasiado, poco, etc.*

## 2.2. Combinaciones

Con respecto a las combinaciones que admite el *sadjn*, el núcleo puede estar antecedido por :  
uno o dos adverbios.

*adv* + (*adv*) + *adj*<sub>1</sub>

uno o dos indefinidos.

*indf* + (*indf*) + *adj*<sub>1</sub>

un adverbio y un indefinido.

*adv* + *indf* + *adj*<sub>1</sub>

tan hermoso

tan increíblemente hermoso

poco amable

bastante poco amable

muy poco amable

## 2.3 El tratamiento del *sadjn* en el paradigma 5P

### 2.3.1 El paradigma 5P

La descripción y formalización se enmarcan en el Paradigma 5P. Las 5P corresponden a:

- P1: P se refiere a Protocolos; un Protocolo representa un dato que ha sido obtenido por un Observador (modelado de modo explícito).

- P2: P se refiere a Propiedades; una Propiedad bajo el punto de vista formal es análoga a un axioma. Un conjunto finito de Propiedades especifica en intensidad un conjunto de secuencias de expresiones de una lengua determinada. Un modelo es una secuencia que satisface un conjunto de Propiedades.

- P3: P se refiere a Proyecciones. Las Proyecciones son generalizaciones sobre las Propiedades o sobre un subconjunto de ellas.

- P4: P se refiere a Principios. Un Principio es una restricción sobre las Proyecciones que son válidas para todas las lenguas o para un conjunto de ellas.

- P5: P se refiere a Procesos. Un Proceso es un procedimiento que está implantado en máquina o que puede ser implantado y con el que se pueden tratar las secuencias de las lenguas naturales.

Se focalizarán las propiedades P2 del Paradigma 5P: propiedades de existencia, de linealidad y de flechado.

### 2.4.2 Propiedades de existencia

Las propiedades determinan:

- cuáles categorías se pueden utilizar en los modelos
- cuáles pueden ser núcleo,
- cuáles aparecen una sola vez

- (iv) cuáles exigen la presencia de otras
- (v) cuáles excluyen la presencia de otras

Debe hacerse referencia a los predicados siguientes:

amod, núcleo, unic, obligdi, exig, excl.

P1: Es la Propiedad que define el vocabulario. Determina las categorías que se pueden utilizar en los modelos y solamente ellas

P1:amod(sadjnesp,[adj, adv, indf]).

El vocabulario está formado por adjetivo, adverbio e indefinido.

P2: Es la Propiedad que determina qué categorías pueden ser núcleo. Una categoría núcleo se anota °cat.

P2 núcleo(sadjnesp, [ °adj]).

Sólo el adjetivo puede ser núcleo.

P3: Es la propiedad que determina cuáles categorías aparecen una sola vez.

P3 unic(sadjnesp, [adj]).

El adjetivo aparece una sola vez.

P4: obligdi(sadjnesp, [adj]).

Esta propiedad expresa que en un sadjn en español hay obligatoriamente un adjetivo.

A continuación, dentro de las propiedades de existencia, se enuncian las de exigencia.

### **Propiedades de exigencia.**

Las propiedades de exigencia expresan que un subconjunto de categorías está exigido en un modelo si se da otro subconjunto de categorías.

P+1:exig(sadjn-esp,[[adv],adj1]]).

*tan hermosa*

P+2:exig(sadjn-esp,[[indf],adj1]]).

*demasiado hermosa*

### **Propiedades de exclusión:**

Las propiedades de exclusión expresan que un subconjunto de categorías está excluido en un modelo si se da otro subconjunto de categorías. Por ejemplo, como se muestra en P 1, el adj 2 va siempre solo, es decir excluye la presencia de otras categorías.

P~1: exclu(snn-esp, [[adj2],[adv]]).

P~2: exclu(snn-esp, [[adj2],[indf]]).

*\*tan internacional*

*\*tan poco internacional*

### **Propiedades de linealidad:**

Expresan las relaciones de orden. Formalmente transforman los conjuntos en listas. La presencia en un modelo de una categoría subsumida por la primera categoría de la fórmula antecede la presencia

de cualquier otra categoría subsumida por otra categoría de la fórmula. Por ejemplo, en P<1 se expresa que el adverbio se antepone al adj 1.

P<1<sadjn-esp,{\_},precede(adv,{adj1})>.

*tan hermosa*

P<2<sadjn-esp{ \_},precede(adv{indf,adj1})>.

*muy poco hermosa*

P<3<sadjn-esp{ \_},precede(indf{adj1})>.

*demasiado hermosa*

### Propiedades de flechado:

Las propiedades de flechado especifican el grafo a partir del cual se calculan las relaciones semánticas. Si varias categorías flechan sobre una (generalmente el núcleo) las que flechan y la que recibe obran como los argumentos de una función a partir de la cual puede calcularse la representación semántica que debe estar asociada estas categorías. Por ejemplo, en F1, adv flecha a adj1

F1 <sadjn-esp,{\_}, flecha(adv, [adj1])>.

*muy hermosa*

F2 <sadjn-esp,{\_}, flecha(indf, [adj.1])>.

*demasiado hermosa*

## 3. Implantación en maquina

A continuación se tratará la implantación en máquina de los sadjn.

### 3.1 Las herramientas

#### Smorph.

Es un segmentador y lematizador, es un software especificado e implementado en el GRIL por Salah Ait-Mokhtar [7], bajo la dirección de Gabriel G. Bès. Es una herramienta totalmente declarativa que trata en una sola etapa todo lo relativo a la presintaxis (tokenización y análisis morfológico), recibe en entrada una secuencia de códigos ASCII y da a la salida una secuencia de ocurrencias asociadas a un conjunto de pares <etiqueta-valor>.

Deben ser declarados los siguientes ficheros que corresponden a cinco tipos distintos de información:

- **Rasgos:** se debe especificar la etiqueta que se desea utilizar, por ejemplo la etiqueta NUM (número) va a tener como rasgos posibles *sg* (singular) y *pl* (plural); GEN (género) que puede ser masculino o femenino; EMS (etiqueta morfo-sintáctica) que contiene los valores adj (que pueden ser tipo1 y tipo2), adv y nom.

- **Terminaciones:** son una serie de caracteres que expresan un rasgo o conjunto de rasgos, por ejemplo *es* para *leales*, *os* para *hermosos*.

- **Códigos ASCII**<sup>5</sup> son una forma de expresar digitalmente el texto, se deben declarar los códigos que se desean utilizar, por ejemplo el espacio, el signo de interrogación, etc.
- **Modelos**: consiste en asociar un conjunto concreto de terminaciones a un conjunto de definiciones morfológicas. Los modelos vienen introducidos por @, que indica el lugar en que va la forma básica o raíz a la que se concatenan las terminaciones.
- **Entradas**: representan el “diccionario lingüístico”, donde se declararán las formas individuales.

## **MPS.**

MPS significa módulo post-smorph. Por medio de reglas que deben ser declaradas, MPS va a trabajar con dos o más estructuras de datos de la salida de Smorph en formato Prolog<sup>6</sup> y va a dar en salida otro formato según el analizador que se vaya a utilizar. MPS ejecuta dos funciones Principales que son la *recomposición* y la *correspondencia*. La recomposición puede ser de dos tipos diferentes: *reagrupamiento* y *división*.

### **Reagrupamiento.**

En la función de reagrupamiento, MPS mediante determinadas reglas puede expresar los valores de entrada (sobre dos o más estructuras de datos de salida de Smorph) y los valores de salida sobre la estructura agrupada.

### **División.**

Las reglas que se declaren con la función de división van a provocar el efecto inverso, a fin de obtener en la salida una secuencia de entidades que sea análoga a las que Smorph asigna a las ocurrencias no contraídas en una cadena. Estas reglas son útiles para tratar las contracciones por ejemplo *del* en español (de él).

### **Correspondencia**

Las reglas que se declaren con la función de correspondencia van a operar sobre una sola estructura de datos a la salida de Smorph y van a poder modificarla en otra estructura de datos. Estas reglas permiten formular en Smorph descripciones básicas, generales y adaptarlas después a la exigencia de cada analizador o de cada aplicación, o enriquecerlas con nuevos pares de *<etiqueta=valor>*.

## **4. ANÁLISIS DEL SADJN (SALIDA DE SMORPH).**

### **Resultados:**

'tan'.

[ 'tan', 'EMS', 'adv' ].

'increíblemente'.

[ 'increíblemente', 'EMS', 'adv' ].

<sup>5</sup> American Standard Code for Information Interchange.

<sup>6</sup> Es un lenguaje de programación lógica. Se usa principalmente en programas de inteligencia artificial y en lingüística computacional, especialmente el procesamiento de lenguaje natural para lo que fue diseñado

'hermoso'.

[ 'hermoso', 'EMS','adj', 'TPOADJ','adj1'].

'naturalmente'.

[ 'naturalmente', 'EMS','adv'].

'desgastado'.

[ 'desgastado', 'EMS','adj', 'TPOADJ','adj1'].

'hermoso'.

[ 'hermoso', 'EMS','adj', 'TPOADJ','adj1'].

## 5. ANÁLISIS DEL SADJN (SALIDA DE MPS).

A partir de los datos obtenidos de las salidas de Smorph, utilizamos la herramienta MPS con el fin de analizar los sadjn. Trabajamos con seis reglas:

**Regla N° 1:** adj da sadjn

*hermoso*

*departamental*

**Regla N° 2:** adv + adj1 da sadjn:

*muy hermosa*

**Regla N° 3:** adv + adv + adj1 da sadjn:

*tan increíblemente hermosa*

**Regla N° 4:** indf + adj1 da sadjn:

*demasiado hermosa*

**Regla N° 5:** indf + indf +adj1

*bastante poco hermosa*

**Regla N° 6:** adv +indf +adj1

*muy poco hermosa*

%1 adj1 da sadjn%

S1 [L1, 'EMS','adj'] --> S1 [L1, 'EMS', 'sadjn'].

% hermosa %



%2 adv+adj1 da sadjn%

S1 [L1, 'EMS','adv'] S2 [L2, 'EMS','adj'] --> S1+S2 [L1+L2, 'EMS', 'sadjn' ].

%muy hermosa %

%3 adv+adv+adj1 da sadjn%

S1 [L1, 'EMS','adv'] S2 [L2, 'EMS','adv'] S3 [L3, 'EMS','adj'] --> S1+S2+S3 [L1+L2+L3, 'EMS', 'sadjn' ].

%tan increíblemente hermosa %

%4indf+adj1 da sadjn%

S1 [L1, 'EMS','indf'] S2 [L2,'adj'] --> S1+S2 [L1+L2, 'EMS', 'sadjn' ].

%demasiado hermosa %

%5 indf+indf+adj1 da sadjn%

S1 [L1, 'EMS','indf'] S2 [L2, 'EMS','indf'] S3 [L3, 'EMS','adj'] --> S1+S2+S3 [L1+L2+L3, 'EMS', 'sadjn' ].

%bastante poco hermosa %

%5 adv+indf+adj1 da sadjn%

S1 [L1, 'EMS','adv'] S2 [L2, 'EMS','indf'] S3 [L3, 'EMS','adj'] --> S1+S2+S3 [L1+L2+L3, 'EMS', 'sadjn' ].

%muy poco hermosa %

'tan increíblemente hermoso'.

[ 'tan increíblemente hermoso', 'EMS', 'sadjn3' ].

'naturalmente desgastado'.

[ 'naturalmente desgastado', 'EMS', 'sadjn2' ].

'hermoso'.

[ 'hermoso', 'EMS', 'sadjn1' ].

%fin de la phrase numero : 1

'demasiado hermoso'.

[ 'tan increíblemente hermoso', 'EMS', 'sadjn4' ].

'bastante poco hermoso'.

[ 'bastante poco hermoso', 'EMS', 'sadjn5' ].

'muy poco hermoso'.

[ 'muy poco hermoso', 'EMS', 'sadjn6' ].

## 6. CONCLUSIONES

El *sadjn* puede definirse como el segmento que empieza con el primer elemento del sintagma adjetivo y finaliza en su núcleo. El *sadjn* lleva siempre como núcleo un adjetivo. Las categorías morfosintácticas que lo integran son: adjetivo, adverbio e indefinido. Hemos clasificado a los adjetivos en dos tipos, tipo 1 (adjetivos que pueden ir solos o anteceditos de adverbios e indefinidos) y tipo 2 (adjetivos que deben ir solos ya que no admiten estas categorías).

La descripción y formalización del *sadjn* se realizó en el marco del paradigma 5P desarrollado por desarrollado en el GRIL (Groupe de Recherche dans les Industries de la Langue) Universidad Blaise-Pascal de Clermont-Ferrand.

Hemos utilizado las herramientas Smorph y MPS para la implantación en máquina.

## Referencias

- [1] Abney, Steven 1991 Parsing by Chunks. En Berwick et al. (1991).
- [2] Aït-Mokhtar, Salah y Chanod, Jean-Pierre Incremental Finite-State Parsing. En Proceedings of the 8th Conference on Applied Natural Language Processing, ANLP-97. Washington, 72-79, 1997
- [3] Gala, Núria Analyse syntaxique automatique avec XIP. Séminaire TAL du DELIC, Universidad de Provence, 2005
- [4] Bès G. G., Hagege, C. Propriétés in 5p. Technical report, Groupe de Recherche dans les industries de la langue (GRIL), URL, 2001.
- [5] Hagege C. Analyse syntaxique automatique du portugais. Tesis de Doctorado GRIL, Univ. Blaise Pascal, 2000
- [6] Hagege C., Bès G. G. Encoding and reusing linguistic information expressed by linguistic Properties. En COLING, Proceedings of the Workshop on Grammar Engineering and Evaluation, Taiwan, 2002
- [7] Aït-Mokhtar S. L'analyse presyntaxique en une seule étape. Tesis de Doctorado, GRIL, Univ. Blaise Pascal, 1998.