

Descripción e implantación en máquina del sintagma adjetivo núcleo (sadjn)

Description and implantation in machine of the nucleus adjective sintagma (sadjn)

Bárbara Méndez

Facultad de Humanidades y Artes

Universidad Nacional de Rosario

Rosario, Argentina

barbaramendez555@hotmail.com

Abstract

This work proposes the description and the implantation in machine of the nucleus adjective sintagma (sadjn). We study how it is composed, its beginning, its ending and its nucleus. From morphosyntactic categories that constitute it (adjectives, adverbs and indefinites), we perform a classification taking as a criterion the positions that they present in sadjn and their possible combinations. For the implantation in machine, we use Smorph and MPS, two computer science tools used for text processing. Smorph performs the morphological analysis and MPS works on local grammars. Both integrate a module, so that the Smorph output is the MPS input.

Keywords: Nucleus adjective sintagma, Description, Implantation in machine.

Resumen

Este trabajo se propone la descripción y la implantación en máquina del sintagma adjetivo núcleo (sadjn). Se estudia su composición, su inicio, su finalización y su núcleo. A partir de las categorías morfosintácticas que lo constituyen (adjetivos, adverbios e indefinidos), realizamos una clasificación tomando como criterio las posiciones que presentan en el sadjn y sus posibles combinaciones. Para la implantación en máquina, nos valemos 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.

Palabras claves: Sintagma adjetivo núcleo, Descripción, Implantación en máquina

1. INTRODUCCIÓN

El objetivo de este trabajo es realizar la descripción y la implantación en máquina del sintagma adjetivo núcleo (sadjn). En lo que respecta a la descripción del sadjn estudiamos como está compuesto, dónde se inicia y dónde termina y cuál es su núcleo. Observamos las categorías morfosintácticas que lo integran (adjetivos, adverbios e indefinidos), y las clasificamos de acuerdo a las posiciones que ocupan en el sadjn y a sus posibles combinaciones. Es necesario mencionar que el adjetivo ya ha sido estudiado en el trabajo “El sintagma adjetivo núcleo” (Méndez, Morosano; Ortiz 2007), por lo tanto estudiamos los adverbios e indefinidos que se presentan en el sadjn. Como consecuencia de esto, nos centramos en los sadjn de dos constituyentes y de tres constituyentes.

Para la implantación, nos valemos 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.

Este trabajo consta de dos secciones. En la sección 1 realizamos la descripción del sadjn, esto comprende estudiar la composición del sadjn, el comienzo, la finalización, el núcleo y las categorías morfosintácticas que lo constituyen. Clasificamos a las mismas y enunciamos todos los casos posibles de sadjn de dos constituyentes y de tres constituyentes. La sección 2 contiene la implantación en máquina, en lo que se refiere a Smorph, presentamos la información declarada y necesaria para la utilización de Smorph y mostramos la salida de este programa utilizando un fragmento del corpus analizado. En lo que respecta a MPS exponemos algunas de las reglas declaradas de sadjn que corresponden al fragmento analizado por Smorph, y damos los resultados obtenidos, es decir la cantidad de sadjn encontrados en el corpus, y el porcentaje de precisión y de cobertura.

2. DESCRIPCIÓN DEL SINTAGMA ADJETIVO NÚCLEO (SADJN)

El sadjn comienza con el primer elemento del sintagma adjetivo y finaliza en su núcleo, por ejemplo en el sintagma adjetivo *muy orgulloso por su triunfo* el sadjn se inicia con el adverbio y termina en su núcleo (*orgulloso*), quedando fuera *por su triunfo*. El sadjn lleva siempre como núcleo un adjetivo (*orgulloso*). Los adjetivos pueden aparecer solos, sin que los anteceda ningún elemento, o pueden ir antecidos por algunos adverbios e indefinidos, como es el caso de *muy orgulloso* y de *poco orgulloso*. Por lo tanto las categorías que integran el sadjn son: adjetivos, adverbios e indefinidos.

De acuerdo con esto, el sadjn puede presentarse con un constituyente (adj.), con dos constituyentes (adv + adj) (indf + adj) y con tres constituyentes (adv+ adv+adj) (adv+indf+adj) (indf+indf+adj) (indf+adv+adj).

2.1. Clasificación de las categorías que integran el sadjn

Como dijimos anteriormente las categorías que integran el sadjn son: adjetivos, adverbios e indefinidos. La clasificación de las categorías nos permite avanzar en la descripción del sadjn.

2.2.1. Adjetivos:

En el trabajo anteriormente mencionado sobre el sadjn (Méndez, Morosano y Ortiz 2007) se clasifica a los **adjetivos** en dos tipos: adj.1 y adj.2.

Adj.1: son los que pueden ir sólo o anteceditos de adverbios e indefinidos. Los participios se comportan como estos adjetivos.

Ej: *atractiva*

más atractiva

demasiado atractiva

Adj.2: deben ir solos

Ej: *internacional*

2.2.2. Adverbios e Indefinidos:

Es importante mencionar que no todos los adverbios e indefinidos integran el sadjn. A continuación presentaremos los adverbios e indefinidos que se presentan en el sadjn.

Adverbios que integran el sadjn:

muy, más, tan, increíblemente (adv terminados en mente)

Indefinidos que integran en el sadjn:

poco, bastante, nada, demasiado, mucho

Otra cuestión fundamental es que los adverbios e indefinidos no presentan las mismas posiciones en el sadjn. Los clasificamos de acuerdo con esto y estudiamos las combinaciones que pueden presentar en el sadjn.

Tabla 1: Clasificación de los adverbios que integran el sadjn

adv1	muy	Va siempre solo <i>muy atractiva</i>
adv 2 adv 2a	tan	Puede ir sólo: <i>tan atractiva</i> Puede ir combinado: tan + poco /adv.mente <i>tan poco atractiva</i> <i>tan increíblemente atractiva</i>
adv 2b	más menos	Puede ir solo <i>más atractiva</i> Puede ir combinado: mucho /increíblemente + más <i>mucho más atractiva</i> <i>increíblemente más atractiva</i>

adv 2c	increíblemente	Puede ir solo <i>increíblemente atractiva</i>
	ciertamente	
	comúnmente	Puede estar combinado: tan + increíblemente
	considerablemente	<i>tan increíblemente atractiva</i>
	constantemente	increíblemente + más/menosi
	esencialmente	<i>increíblemente más atractiva</i>

Tabla 2 : Clasificación de indefinidos que integran el sadjn.

indf.1	demasiado nada	Va siempre solo <i>demasiado atractiva</i>
indf 2 indf2a	bastante	Puede ir solo <i>bastante atractiva</i> Puede ir combinado: bastante + más /menos /poco <i>bastante más atractiva</i> <i>bastante poco atractiva</i>
indf2b	poco	Puede ir sólo <i>poco atractiva</i> Puede ir combinado:tan + poco <i>tan poco atractiva</i>
indf.3	mucho	Nunca va sólo <i>*mucho atractiva</i> Debe ir combinado: mucho+ más / menos <i>mucho más atractiva</i>

2.3. Sadjn con dos constituyentes y con tres constituyentes.

Los adverbios e indefinidos integran el sadjn con dos constituyentes y el sadjn con tres constituyentes. Basándonos en la clasificación realizada podemos observar:

Tabla 3: adverbios e i indefinidos que se presentan en sadjn de dos o de tres constituyentes.

adv 1(muy) indf 1 (demasiado, nada)	Integran sólo el sadjn con dos constituyentes
adv 2a (tan) adv 2b (más, menos) adv 2c (adv.mente) indf 2a (bastante) indf 2 b (poco)	Pueden integrar sadjn con dos constituyentes y con tres constituyentes.
indf 3 (mucho)	Integra sólo sadjn con tres constituyentes

2.3.1. Sadjn con dos constituyentes

Caso 1: adv + adj.1

adv 1a+ adj1	Ej.: <i>muy atractiva</i>
adv 2a + adj.1	Ej.: <i>tan atractiva</i>
adv 2b+adj1	Ej.: <i>más atractiva</i>
adv 2c+adj1	Ej.: <i>increíblemente atractiva</i>

Caso 2: Indf + adj.1

indf 1+adj.1	Ej.: <i>demasiado atractiva</i>
indf 2a+adj.1	Ej.: <i>bastante atractiva</i>
indf 2b+adj.1	Ej.: <i>poco atractiva</i>

2.3.2. Sadjn con tres constituyentes:

Caso 1: adv 2+ adv 2 +adj.1

adv 2a + adv2c	Ej.: <i>tan increíblemente atractiva</i>
adv 2c + adv2b +adj.1	Ej.: <i>increíblemente más atractiva</i>

Caso 2: adv2a + indf2b +adj 1

Ej.: *tan poco atractiva*

Caso 3:indf+ adv +adj1

indf 3+ adv2b +adj1	Ej.: <i>mucho más atractiva</i>
indf 2b+ adv 2b+ adj.1	Ej.: <i>poco más atractiva</i>
indf2a +adv2b +adj1	Ej.: <i>bastante más atractiva</i>

Caso 4: indf +indf+ adj1

indf2a +indf2b +adj.1

Ej.: *bastante poco atractiva*

3. Implantación en máquina

3.1. Smorph

Es un segmentador y lematizador, es un software especificado e implementado en el GRIL por Salah Ait-Mokhtar(1998). 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:

Entradas:

Las entradas representan el diccionario lingüístico (Solana, 2006) Las palabras pueden aparecer directamente con la indicación de los rasgos morfológicos o con la indicación del modelo morfológico al que pertenecen. Tendiendo en cuenta el objetivo de nuestro trabajo, fue necesario clasificar a los adverbios e indefinidos que se presentan en el sadjn tomando como criterio las posiciones que ocupan y sus combinaciones. Así, al tratarse de adverbios e indefinidos que no tienen variaciones morfológicas, se señala su carácter de adverbios y de indefinidos.y en este caso la clasificación correspondiente.

menos /adv2b .

meramente /adv2c .

En el caso de los adjetivos, declaramos el lema, que por convención se expresa con la forma masculina y la indicación del modelo morfológico al que pertenece. Ej.:atractivo @a4 .

Modelos

En los modelos se señala la estructura morfológica. Como puede verse en el ejemplo anterior, @ introduce el modelo e indica el lugar en que va la raíz a la que se concatenan las terminaciones.

En este caso, sólo los adjetivos necesitan de esta información.

@a4 -1

+o adj/masc/sg

+a adj/fem/sg

+os adj/masc/pl

+as adj/fem/pl .

Se indica el número de caracteres que se extrae al lema (atractiv) y se agregan las terminaciones correspondientes. (o, a, os, as)

Terminaciones

Son una serie de caracteres que expresan un rasgo o conjunto de rasgos, por ejemplo *o, a, os, as* para *atractivo, atractiva, atractivos, atractivas*.

Rasgos

Se debe especificar la etiqueta que se desea utilizar, por ejemplo la etiqueta morfosintáctica; EMS que contiene los valores, entre otros, de adj, adv y det.

Se declara la etiqueta morfosintáctica mayor:

EMS	adv	adver
	det	determ
	adj	adje

Como puede verse los indefinidos forman parte de EMS det.

TDET	tipdet	
	art	artic
	dem	demos
	pos	poses
	indf	indef

TADJ	tipadj	
	adj1	adjet1
	adj2	adjet2 .
TINDF	tipindf	
	indf1	indef1
	indf2a	indef2a
TADV	tipadver	
	adv1a	adver1a
	adv2a	adver2a
	adv2b	adver2b
	adv2c	adver2c .

Códigos ASCII

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.

Trabajamos con un corpus de 13.339 palabras, a continuación presentamos una salida de Smorph que corresponde al análisis de un fragmento del corpus analizado.

Apenas si en algunas porciones del patio delantero la tierra parece más floja –los lugares menos transitados-, liberando una capa delgada de arena cuyos cristalitos producen un brillo seco”

'Apenas'.

['apenar', 'EMS','v', 'MODOV','ind', 'PERS','2a', 'NUM','sg', 'TPO','pres', 'TR','r', 'TC','c1'].

'si'.

['si', 'EMS','sub'].

'en'.

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

'algunas'.

['algunas', mi].

'porciones'.

['porción', 'EMS','nom', 'GEN','fem', 'NUM','pl'].

'del'.

['del', 'EMS','contr'].

'patio'.

['patio', 'EMS','nom', 'GEN','masc', 'NUM','sg'].

'delantero'.

['delantero', 'EMS','nom', 'GEN','masc', 'NUM','sg'].

['delantero', 'EMS','adj', 'GEN','masc', 'NUM','sg'].

'la'.

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

['lo', 'EMS','cl', 'TPCL','nrf'].

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

'tierra'.

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

'parece'.

['parecer', 'EMS','v', 'MODOV','ind', 'PERS','3a', 'TPO','pres', 'TR','r', 'TC','c2',].

['parecer', 'EMS','v', 'MODOV','imper', 'PERS','2a', 'TPO','pres', 'TR','r', 'TC','c2',].

'más'.

['más', 'TADV','adv2b'].

'floja'.

['flojo', 'EMS','adj', 'GEN','fem', 'NUM','sg'].

'###'. 1.

'.'.

['gui', 'EMS','guion'].

'los'.

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

['lo', 'EMS','cl', 'TPCL','nrf1'].

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

['los', 'EMS','cl'].

'###'. 2.

'-'.
,

['gui', 'EMS','guion'].

'los'.

['los', 'EMS','cl'].

'lugares'.

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

'menos'.

['menos', 'TADV','adv2b'].

'transitados'.

['transitar', 'EMS','v', 'MODOV','part', 'GEN','masc', 'NUM','pl', 'TR','r', 'TC','c1'].

'-'.
,

['gui', 'EMS','guion'].

','

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

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

['gui', 'EMS','guion'].

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

'liberando'.

['liberar', 'EMS','v', 'MODOV','ger', 'TR','r', 'TC','c1'].

'una'.

['unir', 'EMS','v', 'MODOV','subj', 'PERS','1a', 'NUM','sg', 'TPO','pres', 'TR','r', 'TC','c3'],.

['unir', 'EMS','v', 'MODOV','subj', 'PERS','3a', 'NUM','sg', 'TPO','pres', 'TR','r', 'TC','c3'],.

['unir', 'EMS','v', 'MODOV','imper', 'PERS','3a', 'NUM','sg', 'TPO','pres', 'TR','r', 'TC','c3'].

'capa'.

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

'delgada'.

['delgado', 'EMS','adj', 'GEN','fem', 'NUM','sg'].

'de'.

['de', 'EMS','prep'].

'arena'.

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

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

['arenar', 'EMS','v', 'MODOV','imper', 'PERS','2a', 'NUM','sg', 'TPO','pres', 'TR','r', 'TC','c1'].

'cuyos'.

['cuyos', 'EMS','rel'].

'cristalitos'.

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

'producen'.

['producir', 'EMS','v', 'MODOV','ind', 'PERS','3a', 'NUM','pl', 'TPO','pres', 'TR','irr', 'TC','c3'].

'un'.

['un', mi]

.'brillo'.

['brillo', 'EMS','nom', 'GEN','masc', 'NUM','sg'].

'seco'.

['seco', 'EMS','adj', 'GEN','masc', 'NUM','sg'].

['secar', 'EMS','v', 'MODOV','ind', 'PERS','1a', 'NUM','sg', 'TPO','pres', 'TR','r', 'TC','c1'].

['pf', 'EMS','pun'].

['npf', 'EMS','pun']

3.2. 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 Prolog2 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*.

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.

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).

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>.

A partir de los datos obtenidos de las salidas de Smorph, utilizamos la herramienta MPS con el fin de analizar los sadjn. Trabajamos con 27 reglas que corresponden a todos los casos posibles de sajn de dos y tres constituyentes.

Tabla 4: reglas del sadjn de dos y tres constituyentes.

<p>S1 [L1,'EMS','TADV','adv2a'] S2 [L3,'EMS','adj'] --> S1+S2 [L1+L2, 'EMS','sadjn1'].</p> <p>S1 [L1, 'TADV','adv2c', ''] S2 [L2,'T, 'ADV','adv2b'] S3 [L3,'EMS','adj'] --> S1+S2 [L1+L2+L3, 'EMS','sadjn2'].</p> <p>S1 [L1,'EMS','det', 'TDET','ind2a'] S2 [L2,'TADV','adv2b'] S3 [L3, 'EMS','adj'] --> S1+S2 [L1+L2+L3, 'EMS','sadjn3'].</p> <p>S1 [L1,'EMS','det', 'TDET','ind2b'] S2 [L2,'TADV','adv2b'] S3 [L3,'EMS', 'adj'] --> S1+S2 [L1+L2+L3, 'EMS','sadjn4'].</p> <p>S1 [L1,'EMS','det', 'TDET','indf3'] S2 [L2,'TADV','adv2b'] S3 [L3,'EMS','adj'] --> S1+S2 [L1+L2+L3, 'EMS','sadjn5'].</p>

Aclaraciones:

EMS: etiqueta morfosintáctica; nom: nombre; adj: adjetivo; v: verbo; det:determinante; gen: género, masc:masculino, fem: femenino; num: número; sg: singular; pl:plural; MODOV: modo verbal; ind:indicativo; PERS:persona; TPO:tiempo.

En el fragmento analizado se encontraron dos sadjn, que corresponden a las siguientes reglas.

Regla N°15 : adv2b + adj1 da sadjn:

más atractiva

Regla N° 19: adv.2b + part da sadjn:

más estudiado

3.2.1. Resultados:

'más floja'.

['más flojo', 'EMS', 'sadjn15'].

'menos transitados'.

['menos transitar', 'EMS', 'sadjn19'].

En la tabla 5 se muestran los resultados obtenidos, específicamente la cantidad de sadjn y el porcentaje de precisión y cobertura del corpus.

Tabla 5: Resultados obtenidos

Total de palabras	Total de sadjn	Resultados	
		Cobertura	Precisión
13.339	37	100%	100%

4. CONCLUSIÓN

En este artículo presentamos la descripción y la implantación en máquina del sadjn. Como vimos, el sadjn lleva siempre como núcleo un adjetivo. Si el adjetivo es de tipo 1 puede ir antecedido de uno o de dos adverbios o indefinidos. En cuanto a los adverbios e indefinidos observamos que sólo un número reducido se presenta en el sadjn. Clasificamos tomando como criterio las combinaciones posibles y la posición que pueden ocupar en el sadjn. De acuerdo con esto, la clasificación de estas categorías nos permitió distinguir a los adverbios e indefinidos que pueden aparecer sólo en sadjn de dos constituyentes, de los adverbios e indefinidos que constituyen sadjn de dos o tres constituyentes.

En cuanto a la implantación en máquina, trabajamos con un corpus de 13.339 palabras. A partir de los datos obtenidos de la salida de Smorph utilizamos la herramienta MPS. En función de las exigencias de esta herramienta declaramos 27 reglas que representan todos los casos posibles de sadjn de dos y de tres constiyutentes. El resultado obtenido fue un 100% de precisión y un 100% de cobertura.

Bibliografía:

- [1] Méndez B, Morosano M; Ortiz A. (2007): El sintagma adjetivo núcleo, Revista Infosur
- [2] Aït-Mokhtar S.(1998) L'analyse presyntaxique en une seule étape. Tesis de Doctorado, GRIL, Univ.Blaise Pascal.
- [3] SOLANA, Zulema (2006), Lingüística Informática y enseñanza de lenguas, Rosario, Ediciones Juglaría.