

Formalización e Implantación en Máquina del Sintagma Adjetivo Núcleo (sadjn)

FORMALIZATION AND THE IMPLANTATION IN MACHINE OF NUCLEUS ADJECTIVE SINTAGMA (SADJN)

Bárbara Méndez
GRUPO INFOSUR UNR
Rosario, Argentina
barbaramendez555@hotmail.com

Abstract

This work proposes the description, the formalization based on the properties and the implantation in machine of nucleus adjective sintagma (sadjn). The linguistic description of sadjn involves a study of the categories that constitute it, its beginning, its conclusion and its nucleus. The classification of the categories that comprise the sadjn is conducted. The formalization is based on the properties of the 5P Paradigm, developed in the GRIL (Groupe de Recherche dans les Industries de la Langue) Blaise-Pascal University of Clermont-Ferrand (France) and these linguistic hypotheses are expected to be evaluated through their implantation in a computer system. We use xfst, a finite-state automata, implemented at the Xerox Research Center Europe (XRCE) and the Palo Alto Research Center (PARC)

Key words: Nucleus adjective sintagma, 5P Paradigm, Properties, Evaluation of linguistic hypotheses.

Resumen

En este trabajo se presenta la descripción, la formalización a partir de las propiedades y la implantación en máquina del sintagma adjetivo núcleo (sadjn).

La descripción lingüística del sadjn comprende un estudio de las categorías que lo integran, su inicio, su término y cuál es el su núcleo. Se realiza la clasificación de las categorías que conforman el sadjn. La formalización se realiza a partir de las propiedades del Paradigma 5P, desarrollado en el GRIL (Groupe de Recherche dans les Industries de la Langue) Universidad Blaise-Pascal de Clermont-Ferrand (Francia), y se pretende evaluar estas hipótesis lingüísticas a través de su implantación en máquina. Se recurre a la herramienta xfst, un autómata de estados finitos, implementada en Xerox Research Centre Europe (XRCE) y Palo Alto Reserach Centre (PARC).

Palabras Claves: Sintagma adjetivo núcleo, Paradigma 5P, Propiedades, evaluación de hipótesis lingüísticas.

1. INTRODUCCIÓN

En este artículo nos proponemos realizar la descripción del sintagma adjetivo núcleo (sadjn), su formalización a partir de las propiedades del Paradigma 5P desarrollado en el GRIL (Groupe de Recherche dans les Industries de la Langue) Universidad Blaise-Pascal de Clermont-Ferrand (Francia) [1] y la evaluación de estas hipótesis lingüísticas a través de su implantación en máquina. Se recurre a XFST, un autómata de estados finitos, implementado en Xerox Research Centre Europe (XRCE) y Palo Alto Research Centre (PARC, California, USA). Siguiendo a Trouilleux [2] se presenta una traducción del formalismo Propiedades a expresiones regulares de XFST.

Este trabajo forma parte de mi tesis de doctorado dirigida por la Dra. Zulema Solana: titulada “Análisis automático de textos: el sintagma adjetivo núcleo”.

Cabe agregar que tanto la descripción del sadjn ya ha sido estudiada en trabajos anteriores (Méndez, Morosano, Ortiz [3] y Méndez [4]), de modo que aquí se retoman y desarrollan cuestiones presentes en ambos estudios. Por el contrario, la formalización y la implantación en máquina del sadjn, utilizando la herramienta xfst, representa una nueva etapa en la investigación que estoy realizando.

Este trabajo puede dividirse en dos partes, en primer lugar, realizo la descripción lingüística del sadjn, que implica un estudio de las categorías morfosintácticas que lo conforman, es decir, como está compuesto, dónde se inicia, dónde termina y cuál es su núcleo. El estudio del orden en que se presentan las categorías y de las posibles combinaciones permite plantear hipótesis sobre la configuración del sadjn, determinar y clasificar las categorías con la que será conveniente trabajar. En cuanto a la clasificación, se continúa con la clasificación iniciada en Méndez [4].

En segundo lugar, se lleva a cabo la modelización lingüística del sadjn en el paradigma 5P, expresando las propiedades del formalismo 5P. Luego, se presentan las propiedades escritas como expresiones regulares de XFST

2. DESCRIPCIÓN DEL SINTAGMA ADJETIVO NÚCLEO (SADJN)

Siguiendo a Méndez, Morosano, y Ortiz el sadjn es una construcción que comienza con el primer elemento del sadjn y finaliza en su núcleo, que siempre es un adjetivo. A modo de ejemplo, en el sintagma adjetivo *muy sorprendido por los resultados*, el sadjn se inicia con el adverbio (muy) y termina en su núcleo (sorprendido).

De este modo, el sadjn presenta las siguientes características:

- i) su núcleo es siempre un adjetivo.
- ii) las categorías que lo integran son: adverbios (adv), indefinidos (indf), adjetivos (adj).
- iii) El sadjn puede presentarse: con un constituyente (sorprendido), con dos constituyentes (*muy sorprendido*) con tres constituyente (*tan increíblemente sorprendido*).

3. CLASIFICACIÓN DE LAS CATEGORÍAS QUE INTEGRAN EL SADJN

Como se dijo, nos valdremos de la clasificación de las categorías realizada en Méndez [4].

3.1 Clasificación de adjetivos

En el artículo mencionado se clasifica tomando como criterio que adjetivos admiten ser antecidos por adv e indf. Se distinguen dos tipos de adjetivos:

-adj 1 que pueden ir sólo o antecidos de adverbios e indefinidos

-adj.2 que deben ir solos.

Proponemos aquí otro grupo de adjetivos, que llamaremos adj.3, este grupo no puede ir precedido de adv e indf, tampoco puede ir sólo y siempre modifica al nombre. Además, considerando las posibles combinaciones que presentan los adj.1, esto nos permite distinguir en 2 subtipos: adj1a, adj1b. A continuación presento la clasificación de los adjetivos:

Tabla nº1: Clasificación de los adverbios

Tipos	Características	Ejemplos
<p>Adj 1</p> <p style="text-align: right;">adj1a</p> <p style="text-align: right;">adj1b</p>	<p>-pueden ir solos</p> <p>-Pueden ir combinados de adv e indf</p> <p>antecedidos de : tan, más, muy, adv-en mente, nada, demasiado, bastante, poco, algo.</p> <p>antecedidos además de: bien, mal, casi, recién.</p>	<p><i>alegre, hermoso, atractivo, grande, recordado,</i></p> <p><i>tan alegre, más alegre</i></p> <p><i>bien recordado ,casi recordado</i></p>
<p>Adj2</p>	<p>-van siempre solos</p>	<p>nacional</p>
<p>Adj3</p>	<p>-no pueden ir solos</p> <p>-no pueden combinarse con adverbios e indefinidos</p>	<p><i>*mero</i></p> <p><i>*muy mero</i></p>

3.2 Clasificación de Adverbios:

En lo atañe a los adverbios, el criterio es diferenciar a aquellos adverbios que pueden modificar a los adj1 (adv1), de aquellos que solo pueden modificar a los adj1b (adv2) . De este modo, presentamos la clasificación de los adverbios:

Tabla nº2: Clasificación de adverbios según a que tipo de adjetivo modifica.

Tipos	Características	Ejemplos
Adv1 adv1a (tan) adv1b (más, menos) adv1c (muy) adv1d (adv -mente)	-Pueden combinarse con adj.1	<i>Tan alegre, tan recordado mas alegre , más recordado muy alegre, mas recordado in creiblemente alegre, increiblemente recordado</i>
Adv2 adv2a (bien) adv2b(casi) adv2c (mal) adv2d (recién)	-Pueden combinarse con adj.1b	<i>bien recordado casi recordado mal recordado recien recordado</i>

3.3 Clasificación de indefinidos:

Es importante destacar que para diferenciar a los indefinidos no se utilizó el criterio estudiado en la clasificación de los adverbios ya que los indefinidos pueden modificar a ambos subtipos de adj1. Por esta razón, se estudió la posibilidad de combinación que tienen los indefinidos aquí tratados, indicando el orden en que pueden aparecer en el sadjn. En la tabla nº 3 se diferencian tres tipos de indefinidos.

Tabla nº 3: Clasificación de indefinidos

Tipos	Características	Ejemplos
Indf1 (nada, demasiado, algo)	-Van siempre solos	<i>demasiado atractiva</i>
Indf2 Indf2a (bastante) Indf2b (poco)	-pueden combinarse -va primero y se combina con: mas, bien, mal. - puede ir primero combinado con : más - segundo combinado con: increíblemente	<i>bastante más hermosa, bastante bien recordado.</i> <i>un poco más recordado</i> <i>Increíblemente poco recordado</i>
Indf3	-siempre va combinado con: mucho	<i>mucho más hermosa</i>

3.4 Combinaciones

Teniendo en cuenta la formalización lingüística del *sadjn* que se realizará en la sección siguiente, se estudian las combinaciones de *adv* e *indf*. La tabla nº5 presenta las posibles combinaciones entre: *adv + adv*, *adv + indf*, *indf + indf* e *indf + adv*.

Tabla nº5: Posibles combinaciones entre *adv* e *indf*

	más / menos	adv-mente	mal	bien	poco
tan	-	+	+	+	+
muy	-	-	+	+	-
adv- mente	+	-	+	+	-
bastante	+	-	+	+	+
poco	+	-	-	-	-
mucho	+	-	-	-	-

En esta tabla se considero el orden que pueden presentarse estas categorías. Los adverbios e indefinidos que se presentan en primer lugar se encuentran a la izquierda, en cambio los que están arriba se presentan segundos en la combinación. (algunos se presentan en ambas posiciones).

4. El tratamiento del *sadjn* en el Paradigma 5P

4.1 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.

4.1.1 Propiedades de existencia:

Las propiedades de existencia determinan (i) cuáles son las categorías que se pueden utilizar en los

modelos, (ii) cuáles pueden ser núcleo, (iii) cuáles aparecen una sola vez, (iv) cuáles exigen o excluyen la presencia de otras, (v) cuáles concuerdan entre sí.

Intervienen los predicados siguientes:

amod
núcleo
unic
obligdi
exig
exclu

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.

4.1.2 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. Así, en P+1, en un sadjn un adv1 implica necesariamente la presencia de un adj1.

P+1: exig (sadjn-esp, [[adv1],[°adj1]]). muy hermosa

P+2: exig (sadjn-esp, [[adv2],[°adj1b]]). bien recordado

P+3: exig (sadjn-esp, [[adv2d],[°adj1b]]). recién recordado

P+4: exig (sadjn-esp, [[indf1],[°adj1]]) demasiado hermosa

P+5: exig (sadjn-esp, [[indf2],[°adj1]]) bastante hermosa

P+6: exig (sadjn-esp, [[indf3],[adv1b]]). mucho más (hermosa)

4.1.3 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. La presencia en un modelo de una categoría subsumida por cualquier categoría en la fórmula determina la ausencia en el modelo de una

categoría subsumida por toda otra categoría en la fórmula. Por ejemplo, el adj2, excluye la presencia de adverbios e indefinidos.

P~1: exclu(sadjn-esp, [[°adj2],[adv],[indf]]). *tan internacional

P~2: exclu(sadjn-esp, [[°adj3],[adv],[indf]]). *tan mero

P~3: exclu(sadjn-esp, [[°adj1a],[adv2]]). *mal hermosa

P~5: exclu(sadjn-esp, [[adv1a],[adv1b],[adv1c],[adv2b]]). * tan mas hermosa

P~6: exclu(sadjn-esp, [[adv1a],[indf2a],[indf2c],[indf3]]). * tan bastante hermosa

P~7: exclu(sadjn-esp, [[indf1],[adv]]). * nada muy hermosa

P~8: exclu(sadjn-esp, [[adv1b],[adv1c],[adv2]]). * mas muy hermosa

P~9: exclu(sadjn-esp, [[adv1c],[adv1d],[adv2b]]). * muy increíblemente hermosa

P~10: exclu(sadjn-esp, [[adv1c],[indf2a],[indf2c],[indf3]]). * muy bastante hermosa

P~11: exclu(sadjn-esp, [[adv1d],[indf2a],[indf3]]). * increíblemente bastante hermosa

P~12: exclu(sadjn-esp, [[adv1d],[adv2b]]). * increíblemente casi hermosa

P~13: exclu(sadjn-esp,[[adv2d],[indf]]). * recién nada

P~14: exclu(sadjn-esp,[[adv2d],[adv]]). * recién mas

P~15: exclu(sadjn-esp,[[adv2],[indf1],[indf2b],[indf3]]). * bien algo

P~16: exclu(sadjn-esp,[[adv2a],[adv2b],[adv2c]]). *bien casi

P~17: exclu(sadjn-esp,[[adv2b],[adv2c]]). * casi mal

P~18: exclu(sadjn-esp,[[indf1],[indf]]). * nada bastante

P~19: exclu(sadjn-esp,[[indf2],[indf3]]). * bastante mucho

4.1.4 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. En el primer caso, el adv1 precede al adj.1.

P<1 <sadjn-esp, { }, precede(adv1, { adj1 })>. más hermosa

P<2 <snn-esp, { }, precede(adv2, { adj1b })>. bien recordado

P<3 <snn-esp, { }, precede (indf1, {adj1 })>. demasiado hermosa

P<4 <snn-esp, { }, precede (indf2, {adj1 })>. bastante hermosa

P< 5 <sadjn-esp { },precede(indf3 {adv1b})>. mucho más hermosa

P< 6 <sadjn-esp { },precede(adv1a {adv1d,adv2a, adv2c})>. tan bien recordado

P< 7 <sadjn-esp { },precede(adv1a {indf2b})>. tan poco recordado

P< 8 <sadjn-esp { },precede(adv1d {adv1b, adv2a, adv2c })>. increíblemente más hermosa

P< 9 <sadjn-esp { },precede(adv2b {adv1a})>. casi tan recordado

P< 10 <sadjn-esp { },precede(adv1c {adv2a, adv2c})>. muy bien recordado

P< 11<sadjn-esp{ },precede(indf2a{adv2a, adv2c})>.bastante bien (recordado)

P< 12<sadjn-esp{ },precede(indf2a{indf2b})>.bastante poco (recordado)

P< 13<sadjn-esp{ },precede(indf2a{adv1b})>.bastante más (recordado)

P< 14<sadjn-esp{ },precede(indf2b{adv1b})>.poco más (recordado)

4.2 Implantación en máquina

4.2.1 Xfst

XFST es un autómata de estados finitos, ha sido desarrollado en Xerox Research Centre Europe (XRCE) y Palo Alto Reserach Centre (PARC). Es una interfase interactiva que permite el acceso a algoritmos básicos de cálculo de estados finitos.

Con esta herramienta se pueden implantar las propiedades realizadas anteriormente con el fin de evaluarlas. En XFST primero se definen las categorías, luego se definen las propiedades utilizando los operadores de XFST, y luego con el comando print words se generan los sintagmas que surgen de la compilación de las propiedades detalladas anteriormente. De esta manera, se pueden evaluar las hipótesis lingüísticas al corroborar si lo generado por el autómata es correcto, dado que el autómata las genera a partir de los datos que se proporcionan a partir de la formalización lingüística. Si la formulación lingüística es errónea, los errores se ponen en evidencia en la generación informática de las propiedades y se puede trabajar en interacción perfeccionando la formulación de las hipótesis lingüísticas. A continuación, presentamos en la tabla nº4 la definición de las categorías del sadjn y, en la tabla nº5 un ejemplo de las propiedades del formalismo 5P traducidas a expresiones regulares de XFST.

Tabla nº 4: Definición de las categorías del sadjn

```
define adj1a [ alegre " " ];
define adj1b [ recordado " " ];
define adj1 [ adj1a |adj1b];
define adj2 [ nacional " " ];
define adj [ adj1 |adj2 ];
define adv1a [tan " "];
define adv1b [más " "];
define adv1c [muy " "];
define adv1d [increíblemente " "];
define adv1 [adv1a| adv1b| adv1c| adv1d];
define adv2a [bien " "];
define adv2b [casi " "];
define adv2c [mal " "];
define adv2d [recién " "];
define adv2 [adv2a| adv2b |adv2c| adv2d];
define adv [adv1 | adv2 ];
define indf1 [nada | demasiado | algo " "];
define indf2a [bastante " "];
define indf2b [poco " "];
define indf3 [mucho " "];
define indf2 [indf2a| indf2b ];
define indf [indf1 | indf2 | indf3 ];
define ems [adj|adv|indf ];
regex ems;
```


Tabla nº 5: Propiedades traducidas a expresiones regulares de xfst

```

define amod [ems | % °]+;
define unic1 [ $?adj];
define nucleo [adj ];
define exig1 [[ $[adv1a] & $[adj1]] | ~$[adv1a]];
define exclu1 [ $[adj2] & ~$[adv1a|adv1c||adv1d|adv2a|adv2b]] ~$[adj2]];
define preced1 [adv1a < adj1];
    
```

4.2.2 Resultados

xfst[2]: print words

nacional

poco más alegre

poco más recordado

poco alegre

poco recordado

bien recordado

mal recordado

casi recordado

increíblemente bien recordado

increíblemente mal recordado

increíblemente más alegre

increíblemente más recordado

increíblemente alegre

increíblemente recordado

más alegre

más recordado

tan poco alegre

tan poco recordado

tan bien recordado

tan mal recordado

tan increíblemente alegre

tan increíblemente recordado

tan alegre

tan recordado

alegre
recordado
mucho más alegre
mucho más recordado
muy alegre
muy recordado
demasiado alegre
demasiado recordado
algo alegre
algo recordado
nada alegre
nada recordado
recién recordado
no tan bien recordado
bastante bien recordado
bastante mal recordado
bastante más alegre
bastante más recordado
bastante alegre
bastante recordado

5. CONCLUSIÓN

Hemos realizado la descripción del sintagma adjetivo núcleo (sadjn), su formalización a partir de las propiedades del Paradigma 5P y la evaluación de estas hipótesis lingüísticas a través de su implantación en máquina. Se ha recurrido a la herramienta XFST, realizando la traducción de las Propiedades a expresiones regulares de XFST. Los resultados logrados por el trabajo del autómata corroboran en un 100% las hipótesis lingüísticas que expresan las Propiedades.

Referencias:

- [1] G. G. Bès (1999). La phrase verbale noyau en français. Recherches sur le français parlé, 15:273–358.
- [2] Trouilleux, François 2007. “Specifying Properties of a Language with Regular Expressions”.En TAL
- [3] Méndez B, Morosano M, Ortiz A (2007): Revista Infosur, Nro1, 75-84.
- [4] Méndez B, (2008): Descripción e implantación en máquina del sadjn, Revista Infosur, Nro.2, 105-116.