

# Gramática FTAG del castellano: árboles elementales asociados a los adjetivos

Carrillo Montero, V.      Díaz Madrigal, V.  
Gómez Ojeda, A.

Departamento de Lenguajes y Sistemas Informáticos.  
Facultad de Informática y Estadística.  
Universidad de Sevilla.

e-mail: vjdiaz@obelix.cica.es

## Abstract

Dentro del formalismo FTAG (Feature structure based Tree Adjoining Grammar), se realiza un estudio que permita modelar los adjetivos calificativos en castellano principalmente para la realización de análisis sintácticos. El formalismo FTAG es una adaptación de las gramáticas TAG con el objeto de incluirlas dentro de las gramáticas basadas en unificación.

Palabras claves: **Gramáticas basadas en unificación, TAG, FTAG, Análisis sintáctico.**

## 1 Introducción

Una gramática de adjunción de árboles (Tree Adjoining Grammar, TAG) [Jos75], [JosSha92] consiste en una quintupla  $(\Sigma, N_T, I, A, S)$  donde:

1.  $\Sigma$  es un conjunto finito de símbolos terminales
2.  $N_T$  es un conjunto de símbolos no terminales:  $\Sigma \cap N_T = \emptyset$
3.  $S$  es un símbolo no terminal distinguido:  $S \in N_T$
4.  $I$  es un conjunto finito de árboles finitos, llamados árboles iniciales, y caracterizados de la siguiente forma:

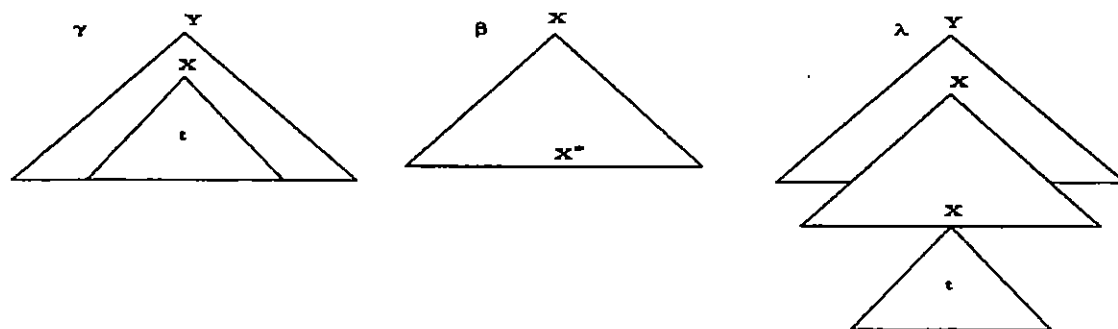


Figure 1: Operación de adjunción

- Los nodos interiores están etiquetados con símbolos no terminales;
  - Los nodos situados en la frontera del árbol están etiquetados mediante símbolos terminales o no terminales. Los símbolos no terminales deben estar marcados mediante  $\downarrow$ , indicando que deben ser sustituidos (Ver más adelante).
5.  $A$  es un conjunto finito de árboles finitos, denominados árboles auxiliares, caracterizados de la siguiente forma:
- Los nodos interiores están etiquetados con símbolos no terminales;
  - Los nodos situados en la frontera del árbol están etiquetados mediante símbolos terminales o no terminales. Los símbolos no terminales salvo uno (nodo pie) deben estar marcados mediante  $\downarrow$ , indicando que deben ser sustituidos. El nodo pie será marcado con  $*$  y su etiqueta debe coincidir con la etiqueta de la raíz.

Los árboles en  $I \cup A$  son denominados árboles elementales. Los árboles elementales son en cierto sentido estructuras lingüísticas mínimas. Las gramáticas TAG, frente a las gramáticas basadas en reglas, poseen un dominio de localidad extendido (árboles de varios niveles). Desde un punto de vista lingüístico este dominio de localidad extendido permite capturar dependencias lejanas [JosVij89] entre las distintas categorías.

Las gramáticas TAG están basadas en un sistema de reescritura de árboles frente a los sistemas de reescritura de cadenas habituales. Los árboles construidos mediante la composición de otros árboles se denominan derivados. En el formalismo TAG se definen dos operaciones de composición: la adjunción y la sustitución.

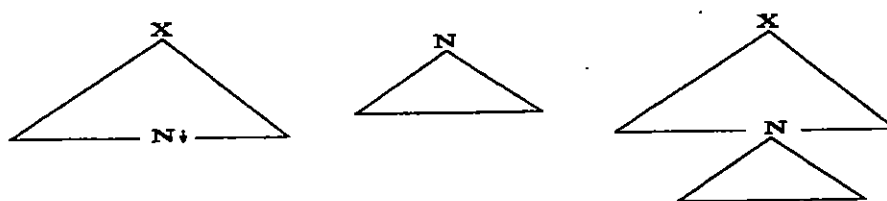


Figure 2: Operación de sustitución

La adjunción (Ver figura 1) construye un nuevo árbol  $\lambda$  a partir de un árbol auxiliar  $\beta$  y otro árbol  $\gamma$  ( $\gamma$  es cualquier árbol inicial, auxiliar o derivado). Sea  $\gamma$  un árbol que contiene un nodo  $n$ , no marcado para sustitución y cuya etiqueta es  $X$ . Sea  $\beta$  un árbol cuya raíz está etiquetada con  $X$ .

El árbol resultante  $\lambda$  de adjuntar  $\beta$  en el nodo  $n$  de  $\gamma$  se obtiene de la siguiente forma. Primero, se poda el subárbol de  $\gamma$  dominado por  $n$ , dejando una copia de  $n$ . Denominemos a este subárbol  $t$ . A continuación, el árbol  $\beta$  se copia sobre el nodo  $n$  identificando su nodo raíz con  $n$ . Por último, el subárbol  $t$  se copia en el nodo pie de  $\beta$  y la raíz de  $t$ , es decir  $n$ , es identificada con el nodo pie de  $\beta$ .

La sustitución (Ver figura 2) tiene lugar sobre un nodo no terminal  $N$  (marcado con  $\downarrow$ ) de la frontera de un árbol. Esta operación consiste en reemplazar el nodo  $n$  por un árbol cuya raíz esté etiquetada con el mismo símbolo que  $n$ .

Como restricciones a estas operaciones, no se permite la adjunción sobre nodos marcados para sustitución y en tales nodos sólo pueden ser sustituidos árboles derivados de árboles iniciales.

En [VijJos88] se define el formalismo FTAG (Feature structure based TAG), consistente en incorporar el formalismo TAG dentro del paradigma de gramáticas basadas en unificación. Para ello se asocian dos estructuras de rasgos a cada nodo  $\eta$  de un árbol elemental: el rasgo inferior  $i_\eta$  y el rasgo superior  $s_\eta$ . Esta doble asociación es provocada por la operación de adjunción sobre nodos.

Si un árbol auxiliar  $\beta$  puede ser adjuntado al nodo  $\eta$  provocará la unificación de  $s_\eta$  con  $s_{raiz}$  y de  $i_\eta$  con  $i_{pie}$  donde *raiz* y *pie* son respectivamente el nodo raíz y pie de  $\beta$ . Al igual que en los formalismos basados en unificación, los rasgos asociados a los nodos permiten capturar restricciones que desde un punto de vista meramente sintáctico son complicadas de capturar.

Las gramáticas TAG han sido estudiadas desde un punto de vista computacional [VijJos85], enmarcándolas dentro de las gramáticas ligeramente sensibles al contexto [JosVijWei91], es decir, están entre la familia de gramáticas CFG y CSG. Actualmente el mejor algoritmo de análisis [ShaJos88] es del orden de  $|G|n^6$ , donde  $|G|$  es una medida del tamaño de la gramática y  $n$  es la longitud de la palabra a procesar. Este analizador está basado en el algoritmo de Early. Existen analizadores más eficientes en los que se aplican restricciones sobre la operación de adjunción.

## 2 Adjetivos y FTAG

Este trabajo está centrado en la definición de los árboles elementales asociados a los adjetivos calificativos con el objeto de poder realizar el análisis de un conjunto amplio de sentencias en lengua castellana. Aquellas composiciones con un componente semántico discutible deberán ser filtradas o interpretadas en una fase adicional.

Siguiendo [Abe91], a la hora de establecer estructuras a los adjetivos no definimos una categoría sintagma adjetival por que consideramos suficiente tener una sólo categoría adjetivo *Adj* que incluirá, si procede, sus posible modificadores. Por tanto, la categoría adjetivo será representada en el formalismo TAG mediante árboles iniciales con raíz etiquetada por *Adj*

Respecto a los rasgos fundamentales asociados a nodos etiquetados *Adj* hemos de tener en cuenta que los adjetivos en castellano, al igual que los sustantivos, poseen género y número a los que se les asociarán respectivamente los rasgos *gen* y *num*. Como sucede en las gramáticas basadas en unificación, estos rasgos intervendrán en ecuaciones de concordancia entre distintas categorías. En las secciones siguientes consideraremos solamente los rasgos inferiores de los nodos.

### 2.1 Grado del adjetivo

Además del género y número, gran parte de los adjetivos, los denominados calificativos, presentan la posibilidad de ser intensificados o comparados, lo que se denomina gradación del adjetivo. Sin embargo, existe un grupo de adjetivos que no admiten modificaciones en su grado, por ejemplo: invisible, vacuno, etc. Debido a estos casos se introduce el rasgo cuantificable *cuant*, que tomará el valor – en los casos anteriores.

Un adjetivo puede encontrarse en grado positivo, superlativo o comparativo. Las estructuras que definiremos están organizadas respecto a la distinta gradación del adjetivo ya que cada grado forma un grupo de construcción característico.

Estos grupos de construcción no pueden ser combinados discrecionalmente, por lo que habilitaremos mecanismos que impidan derivar combinaciones erróneas desde un

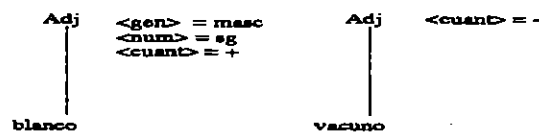


Figure 3: Adjetivo en grado positivo

punto de vista meramente sintáctico. Por ejemplo, no admitimos una intesificación recursiva: tan muy grande como, muy grandísimo, etc.

### 2.1.1 Grado positivo

Los adjetivos en grado positivo señalan una cualidad sin intensificarla ni compararla con otras entidades, por tanto, el rasgo *cuant* asociado a estos adjetivos tiene valor positivo si el adjetivo admite cuantificación. En caso contrario, presenta valor negativo. Se asocian a árboles léxicos cuya raíz transmite el valor del rasgo *cuant* junto con los de número y género (Ver figura 3).

### 2.1.2 Grado superlativo

El grado superlativo señala una cualidad en su más alto grado de relación absoluta o relativa con todas las demás entidades. Los adjetivos en grado superlativo (intensificadores) presentan dos variantes: el superlativo absoluto y el superlativo relativo.

El superlativo absoluto (Ver figura 4 y 5) incluye, por una parte, adjetivos que morfológicamente presentan dicho grado (ej: altísimo) modelados por una estructura similar a la asociada a los adjetivos positivos, pero con rasgo *cuant* = -.

Por otra parte existen otros superlativos absolutos construidos mediante un adverbio o varios, generalmente de modo o cantidad. Estos últimos utilizan un árbol inicial con nodo raíz y nodo a sustituir etiquetados mediante *Adj*. El rasgo *cuant* del nodo raíz tiene valor negativo, mientras que en el adjetivo a sustituir tiene un valor positivo. Con estas definiciones filtramos derivaciones recursivas de superlativos absolutos del tipo: muy altísimo, extremadamente grandísima, etc.

Los superlativos relativos están asociados a un conjunto de árboles iniciales cuyas raíces están etiquetadas con categoría *N*. Incluimos un nuevo rasgo *modif* asociado

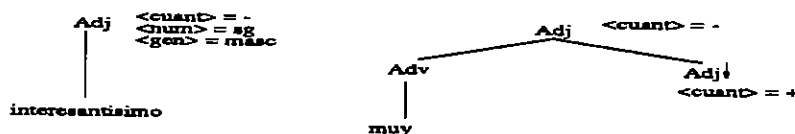


Figure 4: Adjetivo en grado superlativo (1)

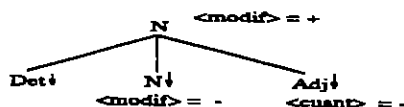


Figure 5: Adjetivo en grado superlativo (2)

a los nodos etiquetados *N*. Dicho rasgo tomará valor positivo en los nodos raíces y negativo en los nodos a sustituir etiquetados *N*. De esta forma evitamos recursiones de superlativos. Además, el nombre y el adjetivo en la frontera del árbol deben coincidir en género y número. Los valores de estos rasgos se transmiten al nodo raíz.

### 2.1.3 Grado comparativo

Los adjetivos en grado comparativo (Ver figura 6 y 7) expresan una cualidad en relación de superioridad, igualdad o inferioridad con la de otras entidades. El rasgo *quant* asociado a la raíz tiene valor negativo, mientras el asociado al adjetivo a sustituir tiene valor positivo. De esta forma se evitan derivaciones recursivas del tipo: tan más grande que como.

Presenta dos estructuras diferentes según que el adjetivo sea sintético o analítico. Los sintéticos son de origen latino y presentan unas estructuras especiales. Los analíticos pueden ser de superioridad, igualdad e inferioridad. Presentan cada uno

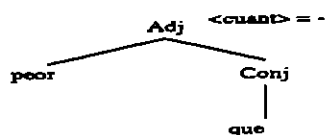


Figure 6: Adjetivo en grado comparativo (1)

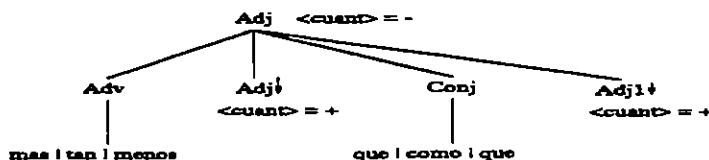


Figure 7: Adjetivo en grado comparativo (2)

de ellos una estructura propia enriquecida de distinta forma según la estructura gramatical a la que pertenece el segundo término de la comparación. Este segundo término puede ser un nombre, un adjetivo, un adverbio, un sintagma preposicional o una sentencia.

## 2.2 Funciones del adjetivo

La categoría *Adj* y sus estructuras asociadas pueden ser combinadas con otras estructuras realizando diversas funciones. Según la función que realice el adjetivo, se presentan particularidades propias que pueden ser modeladas utilizando las características del formalismo gramatical FTAG.

### 2.2.1 Adjunto de nombre

Su función principal es la de referir al sustantivo caracterizándolo o especificándolo semánticamente. Sintácticamente puede tanto anteponerse como posponerse al sustantivo. Algunos adjetivos cambian su significado según se encuentren delante o

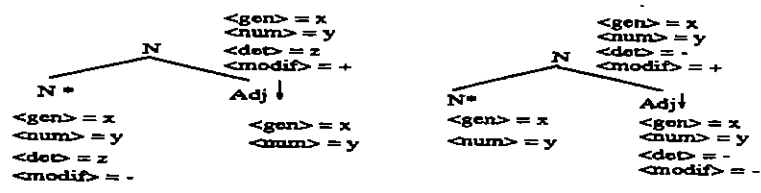


Figure 8: Adjetivo funcionando como adjunto de nombre

detrás del sustantivo que califican. Sin embargo, esta problemática queda fuera de este trabajo debido a que nuestro acercamiento está más centrado en el análisis que en la interpretación.

La estructura (Ver figura 8) en la que aparece será un árbol auxiliar etiquetado con categoría sustantivo *N* en su nodo raíz y pie. La precedencia lineal (LP) se hará explícita mediante dos estructuras diferentes: una que modele la anteposición y otra la posposición.

La recursión explícita en los árboles auxiliares, permitirá una gran variedad de combinaciones de las que deben impedirse formaciones provocadas por adjunciones indiscriminadas de árboles *N* que sitúen mal algún determinante. Esta restricción será impuesta mediante un rasgo *det* que impida dichas adjunciones según tenga valor positivo o negativo.

La concordancia de género y número será explícita mediante sendos rasgos.

### 2.2.2 Atributo

Los atributos (Ver figura 9) se presentan con verbos copulativos y se caracterizan por existir una concordancia de género y número entre sujeto, verbo y adjetivo. La dependencia de estos tres constituyentes es reflejada mediante una única estructura que incluirá la concordancia común de los rasgos *gen* y *num* mediante una ecuación entre los nodos. La estructura asociada será un árbol inicial de raíz *S* dominando las categorías comentadas.

### 2.2.3 Complemento predicativo

Los adjetivos como complemento predicativo están (Ver figura 10) en concordancia en género y número con el sujeto y en número con el verbo predicativo asociado. Consideraremos una estructura *S* donde se recoge el verbo con sus posibles comple-



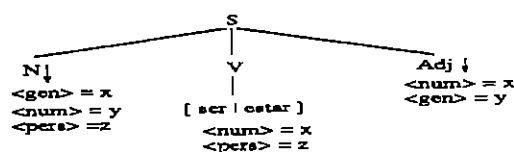


Figure 9: Adjetivo funcionando como atributo

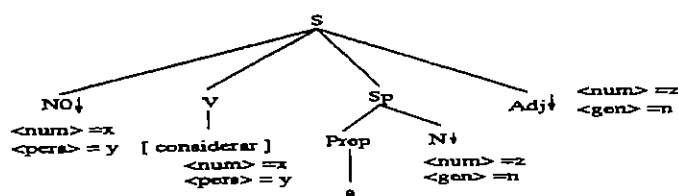


Figure 10: Adjetivo funcionando como complemento predicativo

mentos [Abe91] incluyendo el sujeto. Resolvemos la subcategorización de los verbos mediante estructuras diferentes asociadas a la variedad de combinaciones en los que se involucra el verbo. Este planteamiento es distinto a otros formalismos en los que la subcategorización es portada por un rasgo nuevo.

Esta función del adjetivo estará asociada a una estructura *S* que recoja las distintas relaciones ya comentadas. De manera análoga a la asociada a la función de atributo.

### 2.2.4 Sustantivo

Una característica propia del castellano es la capacidad de sustantivizar los adjetivos mediante el artículo neutro *lo*. Pocos adjetivos admiten sustantivización mediante los artículos usuales. Se modela mediante un árbol (Ver figura 11) inicial con raíz *N*.

Si el adjetivo se encuentra en grado superlativo relativo, se le asocian las estructuras comentadas anteriormente cuando se trató el grado superlativo, con la salvedad de que el nodo *N* a sustituir desaparece (Ver figura 12).

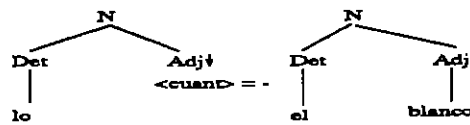


Figure 11: Sustantivación del adjetivo

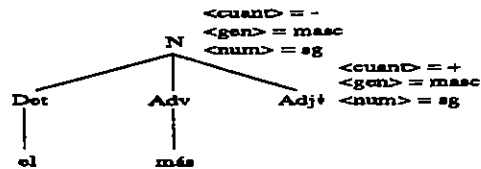


Figure 12: Adjetivo funcionando como sustantivo

### 2.2.5 Adverbio

Algunos adjetivos al ser precedidos mediante preposición funcionan como adverbios. Estos adjetivos tienen asociado un árbol (Ver figura 13) inicial con raíz *Adv*.

Otros adjetivos funcionan como adverbios complementando al verbo. Éstos tienen asociado un árbol auxiliar con etiqueta *V* tanto en su nodo raíz como en su nodo pie.

## 3 Conclusiones

Los adjetivos calificativos son estudiados dentro del formalismo FTAG. Este formalismo incorpora las gramáticas TAG dentro de la familia de gramáticas basadas en unificación. Esta ampliación permite recoger fenómenos lingüísticos difíciles de capturar con el formalismo TAG puro haciendo uso de estructuras de rasgos asociadas a cada nodo de los árboles elementales. Una prueba de ello en este trabajo es que a través de los rasgos *cuant*, *modif* y *det* son filtrados un conjunto de derivaciones



Figure 13: Adjetivo funcionando como adverbio

mal formadas. Además, son definidos un conjunto mínimo de árboles elementales asociados a los adjetivos con el objeto de poder capturar una gran variedad de derivaciones sintácticamente correctas.

## 4 Trabajos futuros

Dentro del marco de las gramáticas FTAG, se completará un conjunto de árboles elementales con el objeto de definir una gramática del castellano que nos permita abordar el análisis de *corpora* y la interpretación semántica.

## 5 Referencias

- Abe91** Abeille, A. (1991). "Une grammaire lexicalisée d'Arbres adjoints pour le Français: Application à l'analyse automatique." Tesis doctoral: Université Paris 7, Paris, France.
- Jos75** Joshy, A. J.; Levy, L., and Takahashi, M. (1975). "Tree Adjunct grammars." *Journal of Computer and System Science*. 10(1) Pags. 136-163.
- JosVij89** Joshy, A. K. y Vijay-Shanker, K. (1989). "Treatment of Long Distance Dependencies in LFG and TAG: Functional Uncertainty in LFG is a Corollary in TAG." *Proceedings 27th Annual Meeting of ACL*. Pags. 220-227.
- JosVijWei91** Joshy, A. K.; Vijay-Shanker, K. y Weir D. (1991). "The Convergence of Mildly Context-Sensitive Grammatical Formalisms." En *Foundational Issues in Natural Language Processing*, ed. por Shieber S. y Wasow T. Pags. 31-82. MIT, Press.

- JosSha92** Joshy, A. K. y Shabes Y. (1992). "Tree Adjoining Grammars and Lexicalized Grammars." En *Tree Automata and Languages*, ed. Nivat M. y Podelsky A. Pags. 409-431. Elsevier Science Publishers B.V.
- ShaJos88** Shabes Y. y Joshy, A. K. (1988). "An Early-Type Parsing Algorithm for Tree Adjoining Grammars." *26th Meeting of ACL.* Pags. 258-269.
- VijJos85** Vijay-Shanker, K. y Joshy, A. K. (1985). "Some Computational Properties of Tree Adjoining Grammars." *3rd Meeting of ACL.*, Pags. 82-93.
- VijJos88** Vijay-Shanker, K. y Joshy, A. K. (1988). "Feature Structure Based Tree Adjoining Grammars." *Proceedings, 12th International Conference on Computational Linguistics*, Pags. 714-719.