

Bases para la desambiguación estructural de árboles de representación sintáctica

Octavio Santana Suárez

Departamento de Informática y Sistemas
Universidad de Las Palmas de Gran
Canaria

osatana@dis.ulpgc.es

Luis J. Losada García

Departamento de Informática y Sistemas
Universidad de Las Palmas de Gran
Canaria

losada@dis.ulpgc.es

José R. Pérez Aguiar

Departamento de Informática y Sistemas
Universidad de Las Palmas de Gran
Canaria

jperez@dis.ulpgc.es

Francisco J. Carreras Riudavets

Departamento de Informática y Sistemas
Universidad de Las Palmas de Gran
Canaria

fcarreras@dis.ulpgc.es

Resumen: Mediante un conjunto de reglas de desambiguación estructural, en este trabajo se reduce el número de árboles de representación sintáctica erróneos que se obtienen por aplicación de las reglas de la gramática del español. Además de ser el punto de partida para un análisis sintáctico automatizado, este proceso complementa los procesos de desambiguación funcional local al elevar su bondad desde un 87% hasta un 96%.

Palabras clave: desambiguación, ambigüedad estructural, lingüística computacional, procesamiento de texto

Abstract: This paper is concerned with the problem of reducing the enormous syntactic representation trees obtained by the application of the spanish grammar rules, by means of a structural deambiguation assembly of rules. This process is the point of departure for an automatic syntax analysis and improves the local functional desambiguation process from 87% to 96%.

Keywords: deambiguation, structural ambiguity, computational linguistic, word processing

1 Las ambigüedades estructurales.

A partir de una combinación de comportamientos funcionales de las palabras de una oración, es posible obtener más de un árbol de análisis cuando se aplica la gramática del español considerada —se formaliza con poco más de 400 reglas—; a tal aparición de múltiples resultados se denomina *ambigüedad estructural*.

La existencia de más de una regla con el mismo símbolo o combinación de símbolos en su parte derecha es lo que se denomina *ambigüedad estructural directa*; en la gramática usada aparecen más de 60

ambigüedades estructurales directas que abarcan unas 240 reglas. Se muestra una de las ambigüedades detectadas:

- *atributo* → *sintagma nominal homogéneo*
- *circunstancial* → *sintagma nominal homogéneo*
- *objeto directo* → *sintagma nominal homogéneo*
- *sujeto* → *sintagma nominal homogéneo*
- *vocativo* → *sintagma nominal homogéneo*

Las ambigüedades estructurales directas originan lo que se llama *conflictos primarios*: aquellas parejas de símbolos no terminales que podrían dar lugar a árboles de análisis erróneos. La tabla 1 detalla los conflictos primarios encontrados.

directo, ni cualquier preposición de un sintagma preposicional un *objeto indirecto*.

Regla: Palabras necesarias

Sea Δ un símbolo no terminal generado a partir de un símbolo Φ y sea $\Gamma(\Delta)$ el conjunto de palabras permitidas para Δ , el símbolo Δ será aceptado si y sólo si entre las palabras que han generado Φ se encuentra alguna del conjunto $\Gamma(\Delta)$.

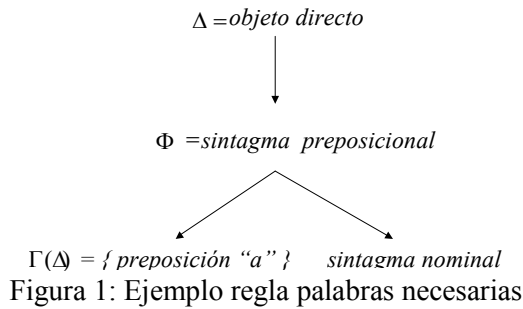


Figura 1: Ejemplo regla palabras necesarias

Es posible aplicar esta regla a diferentes sintagmas:

2.1.1 Los sintagmas preposicionales.

Todo *sintagma preposicional* está constituido por una *preposición* más un *sintagma nominal*. Los diferentes símbolos que presentan conflicto se caracterizan porque no pueden utilizar cualquier preposición: así por ejemplo, el *complemento preposicional del nombre* lleva la preposición “de”.

Δ	Γ																
	a	ante	bajo	con	contr	de	desde	en	entre	hacia	hasta	para	por	según	sin	sobre	tras
<i>atributo</i>						sí	sí					sí					
<i>circ. de afirmación</i>																	
<i>circ. de argumento</i>																sí	
<i>circ. de cantidad</i>	sí					sí	sí										
<i>circ. de causa</i>													sí				
<i>circ. de compañía</i>				sí													
<i>circ. de distancia</i>	sí																
<i>circ. de duda</i>																sí	
<i>circ. de estima</i>						sí	sí										
<i>circ. de fin</i>												sí					
<i>circ. de lugar</i>	sí					sí	sí	sí	sí				sí			sí	
<i>circ. de medio</i>				sí													
<i>circ. de modo</i>				sí												sí	
<i>circ. de oposición</i>					sí												
<i>circ. de tiempo</i>	sí					sí	sí	sí	sí	sí						sí	
<i>compl. adjetivo</i>						sí	sí										
<i>compl. agente</i>						sí						sí					
<i>compl. prepos. del nombre</i>						sí	sí					sí			sí		
<i>objeto directo</i>	sí																
<i>objeto indirecto</i>	sí										sí						
<i>sintagma adverbial</i>	sí					sí											
<i>suplemento</i>	sí					sí	sí					sí					
Conflicto:	Sí	No	No	Sí	No	Sí	Sí	Sí	No	Sí	No	No	Sí	Sí	No	Sí	No

Tabla 2: Las preposicionales en la resolución de conflictos

La tabla 2 muestra qué preposiciones pueden aparecer con cada complemento a que da lugar el sintagma preposicional. En algún caso se eliminan los conflictos —aparecen con un *NO*— y en el resto se reduce la conflictividad. Las preposiciones *cabe* y *so* son anticuadas y actualmente no se usan.

Al analizar “de Juan” en la frase “el coche de Juan es nuevo”, la aplicación de la regla *palabras necesarias* permite rechazar su interpretación como *objeto directo* y mantener la de *complemento preposicional del nombre*.

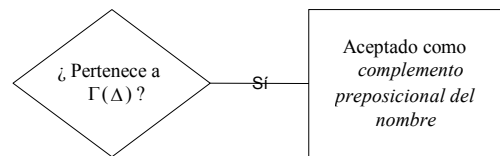
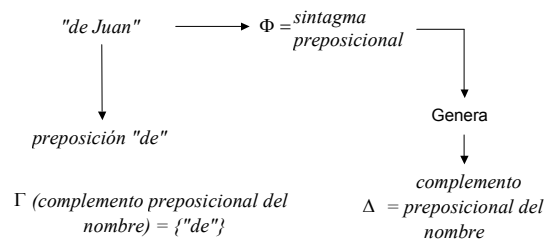


Figura 2: Aplicación de la regla palabras necesarias para el símbolo *complemento preposicional del nombre*

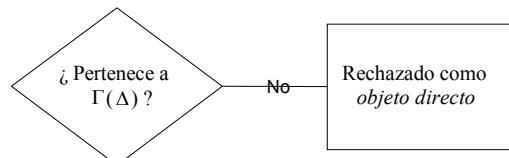
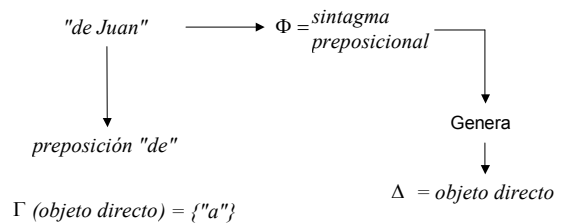


Figura 3: Aplicación de la regla palabras necesarias para el símbolo *objeto directo*

Caso de doble preposición:

Debe tenerse en cuenta que existen casos en los cuales hay una concatenación de preposiciones y es la segunda la que se encuentra asociada al complemento.

“Lo veo **hasta en la sopa**” (complemento circunstancial de lugar)

“Lleva gabardina **hasta en verano**” (complemento circunstancial de tiempo)

“Aquí hay que pedir permiso **hasta para toser**” (complemento circunstancial de finalidad)

“Se mete **hasta con su padre**” (complemento circunstancial de compañía)

Para estos casos, se utiliza la misma tabla de preposiciones para la aplicación de la regla, pero la referencia es la segunda preposición.

Caso de contracción:

El sintagma preposicional puede tener como primer elemento la contracción *al* o *del* que son el equivalente a:

preposición + artículo

Si se separa la contracción en sus dos componentes es posible aplicar las mismas consideraciones de la tabla 2.

Estructura	Contracciones
<i>circunstancial de argumento</i>	<i>del</i>
<i>circunstancial de cantidad</i>	<i>del</i>
<i>circunstancial de distancia</i>	<i>al</i>
<i>circunstancial de lugar</i>	<i>al</i>
<i>circunstancial de precio o estima</i>	<i>del</i>
<i>complemento agente</i>	<i>del</i>
<i>complemento preposicional del nombre</i>	<i>del</i>
<i>objeto directo</i>	<i>al</i>
<i>objeto indirecto</i>	<i>al</i>
<i>suplemento</i>	<i>del</i>

Tabla 3: Las contracciones en la resolución de conflictos

Consideraciones adicionales:

Existe un conjunto de palabras que normalmente no son reconocidas como preposiciones, pero que tienen comportamientos funcionales similares, las llamadas preposiciones imperfectas. Éstas son: *aparte, aun, como, conforme, cuando, durante, donde, mediante, mientras, no obstante, excepto, salvo, más, menos e incluso*. Deben tratarse como preposiciones o no, según corresponda.

Δ	Γ														
	<i>aparte</i>	<i>aun</i>	<i>como</i>	<i>conforme</i>	<i>cuando</i>	<i>durante</i>	<i>donde</i>	<i>mediante</i>	<i>mientras</i>	<i>no</i>	<i>excepto</i>	<i>salvo</i>	<i>menos</i>	<i>más</i>	<i>pero</i>
<i>locución adverbial</i>															<i>si</i>
<i>circ. de afirmación</i>										<i>si</i>					
<i>circ. de argumento</i>															
<i>circ. de cantidad</i>	<i>si</i>										<i>si</i>	<i>si</i>	<i>si</i>	<i>si</i>	
<i>circ. de causa</i>															
<i>circ. de compañía</i>															
<i>circ. de distancia</i>	<i>si</i>					<i>si</i>									
<i>circ. de duda</i>															
<i>circ. de estima</i>															
<i>circ. de fin</i>															
<i>circ. de lugar</i>							<i>si</i>								
<i>circ. de medio</i>	<i>si</i>							<i>si</i>							
<i>circ. de modo</i>			<i>si</i>	<i>si</i>											
<i>circ. de oposición</i>															
<i>circ. de tiempo</i>	<i>si</i>				<i>si</i>	<i>si</i>			<i>si</i>						
<i>compl. agente</i>															
<i>compl. prepos. del nombre</i>															
<i>objeto directo</i>															
<i>objeto indirecto</i>															
<i>suplemento</i>															

Tabla 4: Las preposiciones imperfectas en la resolución de conflictos

2.1.2 Los pronombres personales átonos.

En lo que se refiere a los *pronombres*, la regla *palabras necesarias* puede ser aplicada a los símbolos *objeto directo* y *objeto indirecto*. Tanto el objeto directo como el indirecto pueden estar constituidos por un pronombre personal átono; sin embargo, los pronombres permitidos no son los mismos para ambos.

Estructura	Pronombres
<i>objeto directo</i>	<i>la, las, lo, los, me, nos, os, se, te</i>
<i>objeto indirecto</i>	<i>la, las, le, les, lo, los, me, nos, os, se, te</i>
<i>atributo</i>	<i>lo</i>
<i>morfema de pasiva</i>	<i>se</i>
<i>morfema de impersonal</i>	<i>se</i>
<i>morfema de voz media</i>	<i>me, se</i>

Tabla 5: Los pronombres personales átonos en la resolución de conflictos

2.1.3 Otros casos.

Otros casos para diferentes categorías de palabras son:

Estructura	Categoría	Palabras permitidas
adyacente	adverbio	como, más, menos, no, todo, toda
adyacente	pronombre de relativo	cuya, cuyas, cuyo, cuyos, que
conector subordinante	adverbio	apenas, como, conforme, cuanto, donde, mientras, siempre, tal, tan
conector subordinante	conjunción	aunque, con que, cuando, cuantos, cuantas, para, porque, que, si
construcción comparativa	adverbio	así, como
grupo adjetivo	adverbio	como
grupo adjetivo comparativo	adverbio	más, menos
grupo adjetivo comparativo	otro pronombre	conmigo, contigo, consigo
grupo nominal	adverbio	todo
objeto directo	pronombre de relativo	que, qué
sintagma adverbial	adjetivo	un
subordinada comparativa	conjunción	que
objeto directo	puntuación	dos puntos
presentador	otro pronombre	que, qué

Tabla 6: Otras categorías en la resolución de conflictos

2.2 Las ambigüedades y los símbolos no permitidos.

Se pueden dar casos como el siguiente:

“llegamos tarde nosotros”

“nosotros” puede dar lugar a un *sintagma nominal sustitutivo*, éste a su vez puede dar lugar a un *objeto* o a un *sujeto*. El símbolo *objeto directo* debiera ser rechazado porque los pronombres personales átonos sólo pueden tener función de objeto cuando les precede una preposición.

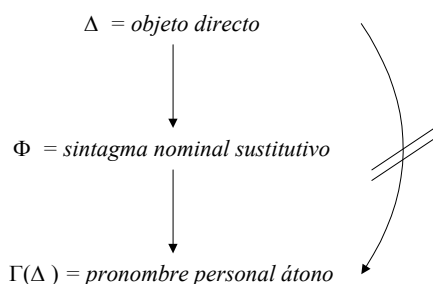


Figura 4: Generación incorrecta de objeto directo

Regla : Símbolos no permitidos.

Sea el símbolo Δ no terminal generado a partir de un símbolo Φ y sea $\Gamma(\Delta)$ un conjunto de símbolos generadores de Φ catalogados como generadores no permitidos de Δ . Entonces Δ generado por Φ será rechazado si Φ ha sido generado por algún símbolo del conjunto $\Gamma(\Delta)$.

Relación de casos

Δ	Φ	$\Gamma(\Delta)$
objeto directo	sintagma nom. sust.	infinitivo
objeto directo	sintagma nom. sust.	otro pronombre
objeto directo	sintagma nom. sust.	pronombre personal átono
objeto indirecto	sintagma nom. sust.	infinitivo
objeto indirecto	sintagma nom. sust.	otro pronombre
objeto indirecto	sintagma nom. sust.	pronombre de relativo
objeto indirecto	sintagma nom. sust.	pronombre personal átono
sujeto	sintagma nom. sust.	pronombre personal tónico
sintagma nom. sust.	núcleo nominal	adjetivo
sintagma nom. sust.	núcleo nominal	infinitivo
sintagma nom. sust.	núcleo nominal	infinitivo con pronombre
sintagma nom. sust.	núcleo nominal	otro pronombre
sintagma nom. sust.	núcleo nominal	pronombre de relativo
sintagma nom. sust.	núcleo nominal	pronombre personal átono
sintagma nom. sust.	núcleo nominal	pronombre personal tónico
sintagma nom. hom.	núcleo nominal	sustantivo

Tabla 7: Símbolos no permitidos

2.3 Las ambigüedades y los símbolos relacionados.

Algunos símbolos no pueden aparecer sin la existencia de otros símbolos en el mismo árbol de análisis.

Regla: Símbolos necesarios.

El símbolo Δ se añade al árbol de análisis si y sólo si existe el símbolo Φ

Relación de casos

- El predicativo obligatorio debe encontrarse acompañado de un núcleo verbal copulativo.
- El objeto directo, el indirecto y otras estructuras que complementan al verbo deben aparecer con un núcleo verbal.

Así, puede definirse el siguiente grupo de relaciones entre estructuras:

Δ - Nuevo símbolo	Φ - Símbolo necesario
atributo	núcleo verbal copulativo
auxiliar pasivo	núcleo verbal pasivo
complemento agente	núcleo verbal pasivo
núcleo verbal copulativo	atributo
núcleo verbal pasivo	auxiliar pasivo
objeto directo	núcleo verbal
oración atributiva	atributo
oración atributiva	núcleo verbal copulativo
oración de suplemento	núcleo verbal
oración de suplemento	suplemento
oración intransitiva	núcleo verbal
oración pasiva	núcleo verbal pasivo
oración transitiva	núcleo verbal
oración transitiva	objeto directo
predicativo objetivo	objeto directo
subordinada comparativa	atributo

Tabla 8: Símbolos necesarios

Consideración de características especiales en los símbolos relacionados:

Para reducir la aparición de objetos directos reconocidos de manera errónea es conveniente tener en cuenta la necesidad de que el verbo sea transitivo.

En la oración “*el niño es alto*”, el *sintagma nominal* “*el niño*” puede dar lugar desde un punto de vista estructural a un *objeto directo*; sin embargo, como “*es*” proviene del verbo “*ser*” que es intransitivo debe rechazarse el símbolo *objeto directo*. El tener en cuenta que los verbos copulativos son a su vez intransitivos ayuda a no confundir un *atributo* con un *objeto directo*. Así pues se redefine la regla anterior como:

Regla: Símbolos necesarios con condición.

El símbolo Δ se añade al árbol de análisis si y sólo si existe el símbolo Φ que cumple la condición $\beta(\Phi)$.

Relación de casos

Símbolo Δ	Símbolos Φ	Condición $\beta(\Phi)$
objeto directo	gerundio	el gerundio debe ser transitivo
objeto directo	infinitivo	el infinitivo debe ser transitivo
objeto directo	infinitivo con pronombre enclítico	el infinitivo con pronombre enclítico debe ser transitivo
objeto directo	participio	el participio debe ser transitivo
objeto directo	verbo	el verbo debe ser transitivo
objeto directo	verbo con pronombre enclítico	el verbo con pronombre enclítico debe ser transitivo

Tabla 9: Símbolos necesarios con condición

2.4 Las ambigüedades y los símbolos incompatibles.

Si se desea que el análisis sea lo más completo posible, se darán casos de reglas del tipo:

- o *oración transitiva* \rightarrow *sujeto* + *predicado*
- o *oración intransitiva* \rightarrow *sujeto* + *predicado*

Como ambas reglas se aplican en una misma situación dan lugar a árboles erróneos. Sin embargo, es posible rechazar el símbolo *oración transitiva* si no se encuentra entre sus nodos inferiores un símbolo *objeto directo*; no obstante, rechazar o aceptar el símbolo *oración intransitiva* es un caso diferente, pues se basa en la no existencia de un símbolo y no en su existencia.

Es por ello que se define el concepto de símbolos incompatibles. De modo que dados dos símbolos incompatibles en un árbol de análisis, éste será rechazado.

Regla: Símbolos incompatibles.

El símbolo Δ se añade al árbol de análisis si y sólo si no existe el símbolo Φ .

Relación de casos

Δ - nuevo símbolo	Φ - símbolo incompatible
atributo	atributo
atributo	morfema de pasiva
atributo	objeto directo
atributo	objeto indirecto
atributo	oración intransitiva
atributo	oración transitiva
atributo	predicativo objetivo
atributo	predicativo subjetivo
atributo	suplemento
complemento agente	suplemento
morfema de impersonal	morfema de impersonal
morfema de impersonal	morfema de pasiva
morfema de impersonal	morfema de voz media
morfema de impersonal	predicativo subjetivo
morfema de impersonal	sujeto
morfema de pasiva	atributo
morfema de pasiva	morfema de impersonal
morfema de pasiva	morfema de pasiva
morfema de pasiva	morfema de voz media
morfema de pasiva	objeto directo
morfema de pasiva	objeto indirecto
morfema de voz media	morfema de impersonal
morfema de voz media	morfema de pasiva
morfema de voz media	morfema de voz media
núcleo verbal	núcleo verbal
núcleo verbal	núcleo verbal copulativo
núcleo verbal	núcleo verbal pasivo
núcleo verbal copulativo	núcleo verbal
núcleo verbal copulativo	núcleo verbal copulativo
núcleo verbal copulativo	núcleo verbal pasivo
núcleo verbal pasivo	núcleo verbal
núcleo verbal pasivo	núcleo verbal copulativo
núcleo verbal pasivo	núcleo verbal pasivo
núcleo verbal pasivo	oración atributiva
núcleo verbal pasivo	oración de suplemento
núcleo verbal pasivo	oración transitiva
objeto directo	atributo
objeto directo	morfema de pasiva
objeto directo	oración de suplemento
objeto directo	oración intransitiva
objeto indirecto	atributo
objeto indirecto	morfema de pasiva
oración atributiva	núcleo verbal pasivo
oración de suplemento	núcleo verbal pasivo
oración de suplemento	objeto directo
oración intransitiva	atributo
oración intransitiva	objeto directo
oración intransitiva	suplemento
oración transitiva	atributo
oración transitiva	núcleo verbal pasivo
predicativo objetivo	atributo
predicativo objetivo	predicativo objetivo
predicativo objetivo	predicativo subjetivo
predicativo subjetivo	atributo
predicativo subjetivo	morfema de impersonal
predicativo subjetivo	predicativo objetivo
predicativo subjetivo	predicativo subjetivo
sujeto	morfema de impersonal
suplemento	atributo
suplemento	complemento agente
suplemento	oración intransitiva
suplemento	suplemento

Tabla 10: Símbolos incompatibles

En este grupo se encuentran aquellos elementos que no se pueden duplicar dentro de un predicado —salvo en oraciones compuestas, en cuyo caso se trataría de varios predicados.

2.5 Las concordancias.

Entre las diferentes estructuras que componen una oración se dan casos de obligatoriedad en la concordancia de ciertas características.

Regla : Concordancias.

Sean los símbolos Δ y Φ de un árbol de análisis, este árbol es aceptado si y sólo si existe concordancia en el conjunto de características definidas para esos símbolos, $\Gamma(\Delta, \Phi)$

Relación de casos

Existe un conjunto de concordancias resueltas durante el proceso de desambiguación funcional. Se caracterizan por atender a parejas ordenadas de símbolos adyacentes. Las concordancias deben ser de nuevo verificadas durante el proceso de creación de los sintagmas, ya que pueden aceptarse parejas ante la posibilidad de que la unión constituya los límites entre estructuras de carácter local, aunque no deben aceptarse como parte de un mismo sintagma. A continuación se detallan las concordancias analizadas durante la desambiguación funcional.

- adjetivo + adjetivo: concordancia en género y número
- adjetivo + adjetivo demostrativo: concordancia en género y número
- adjetivo + adjetivo posesivo antepuesto o pospuesto: concordancia en género y número
- adjetivo + adjetivo posesivo pospuesto: concordancia en género y número
- adjetivo + infinitivo: el adjetivo debe encontrarse en masculino singular
- adjetivo + sustantivo: concordancia en género y número
- adjetivo demostrativo + adjetivo: concordancia en género y número
- adjetivo demostrativo + infinitivo: el adjetivo debe encontrarse en masculino singular
- adjetivo demostrativo + sustantivo: concordancia en género y número
- adjetivo posesivo antepuesto + adjetivo: concordancia en género y número
- adjetivo posesivo antepuesto + infinitivo: el adjetivo debe encontrarse en masculino singular
- adjetivo posesivo antepuesto + sustantivo: concordancia en género y número
- adjetivo posesivo antepuesto o pospuesto + adjetivo: concordancia en género y número
- adjetivo posesivo antepuesto o pospuesto + infinitivo: el adjetivo debe encontrarse en masculino singular
- adjetivo posesivo antepuesto o pospuesto + sustantivo: concordancia en género y número
- artículo determinado + adjetivo: concordancia en género y número
- artículo determinado + infinitivo: el artículo debe encontrarse en masculino singular
- artículo determinado + pronombre de relativo: concordancia en género y número
- artículo determinado + sustantivo: concordancia en género y número
- artículo indeterminado + adjetivo: concordancia en género y número
- artículo indeterminado + infinitivo: el artículo debe encontrarse en masculino singular
- artículo indeterminado + pronombre de relativo: concordancia en género y número
- artículo indeterminado + sustantivo: concordancia en género y número
- forma verbal personal + participio: el verbo que precede al participio es *haber* o *ser*
- infinitivo + adjetivo: el adjetivo debe encontrarse en masculino singular
- infinitivo + adjetivo demostrativo: el adjetivo debe encontrarse en masculino singular
- infinitivo + adjetivo posesivo antepuesto o pospuesto: el adjetivo debe encontrarse en masculino singular
- infinitivo + adjetivo posesivo pospuesto: el adjetivo debe encontrarse en masculino singular
- preposición + preposición: la primera preposición debe ser *a* o *hasta*
- pronombre de relativo + adjetivo posesivo antepuesto o pospuesto: concordancia en género y número
- pronombre de relativo + adjetivo posesivo pospuesto: concordancia en género y número
- sustantivo + adjetivo: concordancia en género y número
- sustantivo + adjetivo demostrativo: concordancia en género y número
- sustantivo + adjetivo posesivo antepuesto o pospuesto: concordancia en género y número
- sustantivo + adjetivo posesivo pospuesto: concordancia en género y número

Concordancia entre *sujeto* y *núcleo verbal*: en las oraciones con verbo en forma personal debe concordar en número y persona con el núcleo de la estructura del sujeto.

$$\Gamma(\text{sujeto}, \text{núcleo verbal}) = \{\text{número}, \text{persona}\}$$

$$\Gamma(\text{sujeto}, \text{núcleo verbal pasivo}) = \{\text{número}, \text{persona}\}$$

$$\Gamma(\text{sujeto}, \text{núcleo verbal copulativo}) = \{\text{número}, \text{persona}\}$$

Además de la concordancia entre sujeto y predicado, se deben cumplir las concordancias entre los siguientes descendientes del predicado —siempre en género y número—:

- *Adyacente* con *núcleo nominal*.
- *Objeto directo* con *objeto directo*.
- *Objeto indirecto* con *objeto indirecto*.
- *Predicativo objetivo* con *objeto directo*.
- *Predicativo subjetivo* con *núcleo verbal*.
- *Predicativo subjetivo* con *sujeto*.
- *Presentador* con *núcleo nominal*.

Consideraciones adicionales

Desde el punto de vista de la implantación de las reglas de concordancia, se hacen las siguientes consideraciones respecto a las flexiones de algunas palabras:

- Los adverbios *más*, *menos* y *no* se considera que actúan con las flexiones masculino, femenino, singular y plural.
- A los artículos determinados *lo* y *los* se les aplica la flexión masculino.
- A los sustantivos se les aplica la flexión tercera persona. Se considera la flexión plural para los sustantivos *gente*, *todo*, *infinidad*, *resto*, *millar*, *caterva*, *montón*, *tipo*, *multitud* y *vida*. Si el sustantivo no tiene asociadas las flexiones de género o número se le asignan todas las posibles. Cuando los numerales representan las horas del día, se les debe asignar la flexión plural y el género femenino.
- A todo adjetivo se le imputa la flexión tercera persona. A los adjetivos numerales a partir del dos se les dota de la flexión plural. A *mal* se le asigna masculino y singular; a *cada* masculino, femenino y singular.
- Para los pronombres de relativo *qué* y *que* se mantienen las flexiones masculino, femenino, singular, plural y tercera persona.
- Para los pronombres personales *yo*, *nosotros*, *nosotras* la flexión debe ser primera persona; para *tú*, *vosotros*, *vosotras*, *usted*, *ustedes* segunda persona; para *él*, *usted*, *ustedes*, *ella*, *ellos*, *ellas* tercera persona; para *me* masculino, femenino y primera persona; para *te* masculino, femenino, y segunda persona; para *se* masculino, femenino y tercera persona; para *la*, *lo* y *le* tercera persona.
- A las palabras que entran en el grupo *otro pronombre* se les asigna la flexión tercera

persona; a *eso* masculino; plural a *aquello*; femenino a *esa* y *aquella*; a *nadie*, a *algo*, a *todo* y a *toda* singular; plural a *todos* y *todas*.

- Para los infinitivos se toma la flexión singular y tercera persona.
- Si tienen como flexión primera persona, a las palabras con categoría funcional *verbo* y *verbo con pronombre enclítico* se les asigna también tercera persona.

2.6 Información semántica.

Cabe introducir información semántica para la eliminación de ambigüedades mediante el análisis del contenido de las palabras —se parte de la información proporcionada por diccionarios ideológicos. Para ello, será necesario introducir una nueva regla de análisis.

Regla: Semántica necesaria.

Sea un conjunto de palabras Ω que se unen para formar el símbolo Δ y sea $\Pi(\Delta)$ el conjunto de significados ideológicos asociados al símbolo Δ , entonces se rechaza Δ si no existe ninguna palabra en Ω tal que su análisis ideológico pertenezca a $\Pi(\Delta)$

En el ejemplo de la figura 5, el *complemento circunstancial de tiempo* se rechaza porque entre las palabras que lo forman ninguna indica tiempo o momento.

Del análisis ideológico de *campesino* resulta que ninguna de sus ideas relacionadas tiene que ver con tiempo o momento.

Modos de vida		
Grupos sociales	Hábitat	Actividades
Nacionalidad	Campo	Agricultor

Tabla 11: Análisis ideológico de *campesino*

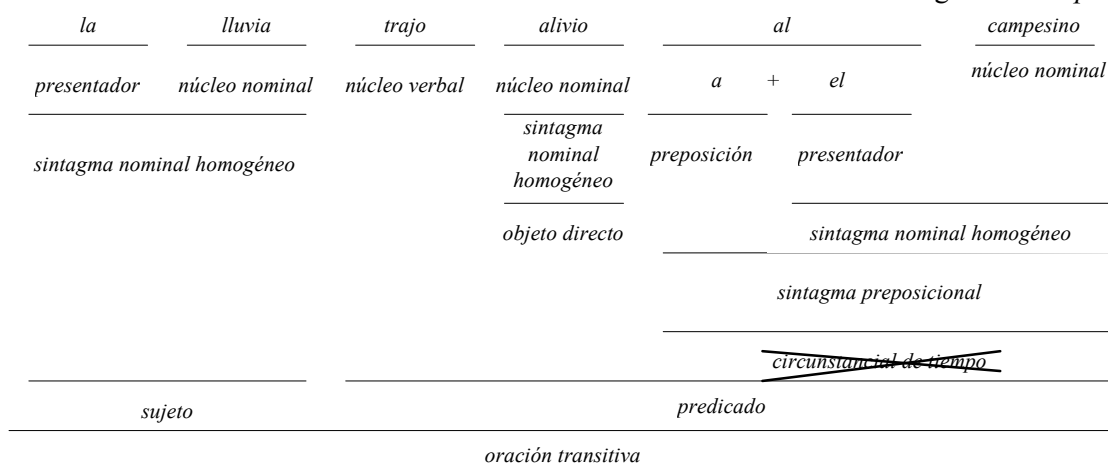


Figura 5: Circunstancial de tiempo rechazado

Se presenta a continuación un ejemplo con un *circunstancial de tiempo*.

“El **domingo** os cazaron y pescaron estas perdices y corvinas”

La palabra núcleo del *circunstancial de tiempo* es *domingo* y aparece en el diccionario ideológico dentro del grupo de *segundo*, como medida de *cantidad*; ya que la cantidad es de tiempo, puede aceptarse el análisis.

Relaciones generales
Cantidad
Segundo

Tabla 12: Análisis ideológico de *domingo*

Es necesario disponer para cada símbolo Δ el conjunto de significados ideológicos aceptables $\Pi(\Delta)$.

Δ	$\Pi(\Delta)$
<i>circ. de tiempo</i>	Relaciones generales – Cantidad – Segundo
<i>circ. de tiempo</i>	Relaciones generales – Tiempo – Tiempo
<i>circ. de tiempo</i>	Relaciones generales – Tiempo – Periodo

Tabla 13: Análisis ideológicos asociados al *circunstancial de tiempo*

Ejemplos:

“**Siempre** algún forastero se presentaba en las augustas salas”

Relaciones generales	El individuo	Modos de vida
Tiempo	Conocimiento	Comunicación
Frecuencia	Certeza	Preposición

Tabla 14: Análisis ideológico de *siempre*

hablaron	los	representantes	de	los	municipios	Nivel 1
				artículo	sustantivo	Nivel 2
			preposición	presentador	núcleo nominal	Nivel 3
			preposición subordinante	sintagma nominal homogéneo		Nivel 4
	artículo	sustantivo	sintagma preposicional			Nivel 5
verbo	presentador	núcleo nominal	complemento preposicional del nombre			Nivel 6
forma verbal simple	sintagma nominal heterogéneo					Nivel 7
núcleo verbal	circunstancial de lugar					Nivel 8
	predicado					Nivel 9
	oración intransitiva					Nivel 10

Figura 6: Aceptación incorrecta del *circunstancial*

“Los gatos como el perro, permanecieron el **día** encerrados”

Relaciones generales	El individuo		Naturaleza
Tiempo	Creencia	Conducta	Atmósfera
Edad	Ritual	Tradición	Clima

Tabla 15: Análisis ideológico de *día*

“Luis y él pasaron el **verano** en Mendoza”

Relaciones generales
Cantidad
Segundo

Tabla 16: Análisis ideológico de *verano*

Sería conveniente crear una nueva disposición ideológica que contuviera las palabras necesarias para la eliminación de este tipo de ambigüedades; los grupos serían uno por cada complemento circunstancial y no varios como ocurre si se utiliza un diccionario ideológico clásico. Con ello se lograría una mejora en la eficiencia de la automatización del proceso de análisis sintáctico.

Consideraciones

Para la oración “hablaron los representantes de los municipios”, es posible que “los representantes de los municipios” sea reconocido como complemento circunstancial de lugar, figura 6. Esto se debe a que la palabra “municipio” figura asociada a lugar en los diccionarios ideológicos; sin embargo, para este caso, el problema es fácilmente resoluble si se atiende sólo a las palabras con categoría funcional en el nivel más alto del supuesto

en	el	granero	Nivel 1
	artículo	sustantivo	Nivel 2
preposición	presentador	núcleo nominal	Nivel 3
preposición subordinante	sintagma nominal homogéneo		Nivel 4
sintagma preposicional			Nivel 5
complemento preposicional del nombre			Nivel 6

Figura 7: Niveles con palabras relevantes

complemento circunstancial de lugar: “los representantes”.

Si se tomara como ejemplo “*Blancas palomas picoteaban en el granero*”, para el complemento circunstancial de lugar “*en el granero*” se obtiene el análisis de la figura 7. De tener en cuenta el nivel superior, se analizaría solamente la palabra irrelevante —*presentadores, conectores y preposiciones*—; por ello, hay que llegar hasta el nivel más alto que contenga palabras relevantes.

2.6.1- Relaciones ideológicas y símbolos.

En la oración “*el niño compró el libro*” podría darse el caso de que “*el niño*” fuera reconocido como objeto directo, figura 8; sin embargo, al tratarse de una persona —información ideológica—, sólo podría ser *objeto directo* si formara parte de un *sintagma preposicional*.

Regla: Semántica incompatible.

Sea un conjunto de palabras Ω que se

el	niño	compró	el	libro
artículo	sustantivo	verbo	artículo	sustantivo
presentador	núcleo nominal	forma verbal simple	presentador	núcleo nominal
sintagma nominal homogéneo		núcleo verbal	sintagma nominal homogéneo	
objeto directo			
predicado				
oración transitiva				

Figura 8: Reconocimiento incorrecto de un objeto directo

unen para formar el símbolo Δ a partir del símbolo Γ y sea $\Pi(\Delta, \Gamma)$ el conjunto de significados ideológicos que producen el rechazo del símbolo Δ , generado a partir de Γ , entonces se rechaza Δ si existe alguna palabra en Ω tal que su análisis ideológico pertenezca a $\Pi(\Delta, \Gamma)$

2.6.2- Relaciones ideológicas entre símbolos.

Existen casos en los cuales puede aplicarse un análisis de las relaciones de tipo ideológico entre los símbolos de una oración. Sea la oración “*el niño compró el libro*”, desde un punto de vista puramente estructural tanto “*el niño*” como “*el libro*” podrían realizar la función de sujeto pues son sintagmas nominales que concuerdan con el verbo: se hace necesario buscar relaciones de tipo semántico. El núcleo verbal es un verbo que implica una acción —*comprar*— y es posible asociar a este tipo de verbos —sólo cuando se encuentran en modo activo— con sujetos que

el	niño	compró	el	libro
<i>artículo</i>	<i>sustantivo</i>	<i>verbo</i>	<i>artículo</i>	<i>sustantivo</i>
<i>presentador</i>	<i>núcleo nominal</i>	<i>forma verbal simple</i>	<i>presentador</i>	<i>núcleo nominal</i>
<i>sintagma nominal homogéneo</i>		<i>núcleo verbal</i>	<i>sintagma nominal homogéneo</i>	
...			objeto	
<i>predicado</i>				
<i>oración transitiva</i>				

Figura 9: Reconocimiento incorrecto de sujeto

sean seres vivos: se aceptaría “*el niño*” y no “*el libro*” como sujeto, figura 9.

Regla : Relaciones ideológicas entre símbolos.

Sean los símbolos Δ y Φ de un árbol de análisis, este árbol es aceptado si y sólo si se cumple la concordancia ideológica en el conjunto de características definidas para esos símbolos, $\Gamma(\Delta, \Phi)$

Esta regla se aplica al *núcleo del sujeto* con el *núcleo verbal*.

2.7 Casos especiales.

En ocasiones, la aplicación de los métodos anteriores no resuelve el problema: se deben resolver como casos especiales.

2.7.1 Propositiones.

La *proposición* y la *proposición subordinada* se diferencian en cómo se unen al resto de la oración —la *proposición* requiere un conector que actúa como coordinante y la *proposición subordinada* de un conector subordinante.

La *proposición de infinitivo* debe tener como *núcleo verbal* un *infinitivo*.

2.7.2 Oraciones interrogativas y exclamativas.

Las oraciones interrogativas y exclamativas tienen la misma estructura y se diferencian en los signos de puntuación que las delimitan: las exclamaciones para la *oración exclamativa* y las interrogaciones para la *oración interrogativa*.

2.7.3 Doble objeto directo.

Puede aparecer un doble objeto directo —dislocación izquierda. Es fácilmente reconocible por las siguientes características:

- 1) Se encuentran el uno junto al otro.
- 2) El primero se encuentra al principio de la oración
- 3) El segundo es un clítico pronominal
- 4) Debe existir concordancia en género y número entre los núcleos del objeto directo.

A Juan lo vi por primera vez el año pasado.

Ese jarrón lo compré el año pasado.

El *doble objeto directo* tiene la siguiente estructura:

objeto directo + objeto directo + verbo

Donde el segundo objeto directo está formado por un *sintagma nominal sustitutivo*.

Regla: Doble objeto directo

Sea Δ un símbolo raíz que cubre toda la oración, si entre sus descendientes aparecen dos símbolos *objeto directo*, entonces Δ es aceptado si y sólo si los *objetos directos* son adyacentes, van seguidos por un *núcleo verbal* y el segundo *objeto directo* está constituido por un *pronombre personal tónico*.

En la oración “*el niño estudia la lección*”, se podría tomar por error un doble objeto directo ya que “*la*” como sustantivo daría

lugar a un *sintagma nominal* que produciría el primer *objeto directo* y “lección” formaría el segundo.

Debe tenerse en cuenta que en oraciones compuestas pueden aparecer dos *objetos directos* por cada *forma verbal personal*: cada pareja debe cumplir la condición de doble objeto directo.

2.7.4 Eliminación de opciones según la posición de los presentadores.

a) Caso de presentadores pospuestos.

Cuando el símbolo *sintagma nominal homogéneo* contiene una subestructura del tipo

“núcleo nominal + presentador”

el *presentador* nunca podrá ser un *adjetivo posesivo antepuesto*, ya que debería ir delante del *núcleo nominal*.

Para la oración “besó Juan su mano” se puede generar la siguiente respuesta:

- Juan: sustantivo
- su: adjetivo posesivo antepuesto
- mano: sustantivo

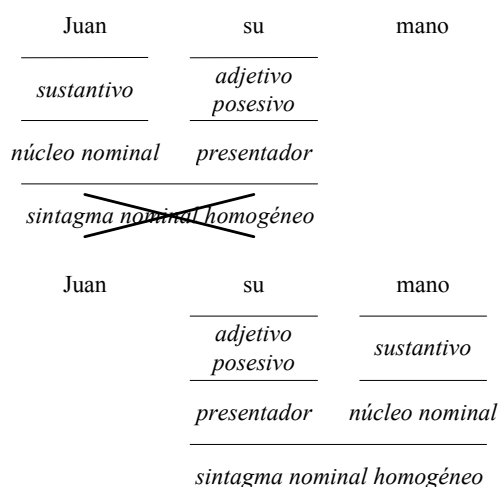


Figura 10: Presentador antepuesto

Podría generar un símbolo *sintagma nominal homogéneo* con:

“Juan su” \Leftrightarrow “núcleo nominal + presentador”

Pero quedaría rechazado por la siguiente regla de *presentadores pospuestos*.

Regla: Presentadores pospuestos

Sea Δ un símbolo que pertenece al grupo de los *sintagmas nominales* y que se

genera a partir de una secuencia de símbolos donde aparece la subsecuencia “núcleo nominal + presentador”, Δ será aceptado si y sólo si el símbolo *presentador* se encuentra en el grupo Ω de los símbolos terminales que pueden ir tras el *núcleo nominal*.

b) Caso de presentadores antepuestos.

El *sintagma adjetivo* puede encontrarse tras un *presentador* cuando el *presentador* finaliza un *sintagma nominal* o ante un *sustantivo* cuando el *sustantivo* abre un *sintagma nominal*. Esto da lugar a que se den casos en los que se generen *sintagmas adjetivos* de modo erróneo: así en “el rojo coche” se da una secuencia del tipo

presentador + adjetivo + sustantivo,

que podría generar un símbolo *sintagma adjetivo* a partir del *adjetivo*, aunque la función real del *adjetivo* es la de *adyacente*. Para que nunca se genere un *sintagma adjetivo* tras un *presentador antepuesto*, es necesario saber qué *presentadores* deben ir antepuestos.

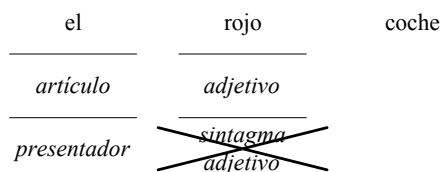


Figura 11: Rechazo del sintagma adjetivo

Regla: Presentadores antepuestos

Sea Δ el símbolo *sintagma adjetivo* que se genera a partir de un símbolo *adjetivo*, Δ no será aceptado si se encuentra tras un símbolo *presentador* del grupo Ω de los símbolos terminales que no pueden ir tras el *núcleo nominal*.

Relación de casos

Los símbolos *adjetivo demostrativo*, *artículo determinado* y *otro pronombre* sólo aparecen antepuestos al *núcleo nominal*. Los adjetivos posesivos pueden dividirse en: aquellos que sólo pueden preceder al núcleo nominal —*mi, mis, tu, tus, su, sus*—, aquellos que sólo pueden ir pospuestos —*mío, mía, míos, mías, tuyo, tuya, tuyos, tuyas, suyo, suya, suyos, suyas*— y los que pueden figurar tanto antepuestos como pospuestos —*nuestro, nuestra, nuestros, nuestras, vuestro, vuestra, vuestros, vuestras*.

2.7.5 Conectores.

Existe un conjunto de combinaciones de palabras que dan lugar a conjunciones conjuntivas; se consideran las siguientes: *a consecuencia, a distinción de, a fin de, a fin de que, a lo que parece, a medida que, a menos que, a pesar, a pesar de, ahora bien, ahora que, al menos, al objeto de, al objeto de que, al parecer, al paso que, antes bien, así como, así es que, así pues, así y todo, aun cuanto, aunque más, bien así como, bien que mal, cada cuanto que, cada y cuanto que, como quiera que, como si, con objeto de, con objeto de que, con todo, con todo eso, con todo esto, conocer se conoce que, cuanto más, cuanto más que, cuanto y más que, dado que, de consiguiente, de forma que, de manera que, de modo que, de suerte que, delante que, en consecuencia, en demasía, en inteligencia de que, en la inteligencia de que, en manera que, entre que, más que, mientras más, mientras que, no obstante, o bien como, para que, pese a, por consecuencia, por consiguiente, por donde, por el consiguiente, por lo menos, por lo mismo, por manera que, por mucho que, por que, presupuesto que, pues bien, pues que, puesto que, que hasta, según y como, según y conforme, si bien, si ya, siempre que, siempre y cuando que, supuesto que, tan pronto, tan siquiera, tan y mientras, visto que, ya que.*

2.7.6 Otros casos.

Existen determinadas situaciones en las que pueden resolverse ambigüedades a partir de consideraciones sobre las palabras, categorías gramaticales y objetos que intervienen:

- ◆ Los **adverbios *sí* y *no***. Las palabras *sí* y *no* cuando actúan como adverbios dan lugar a un sintagma adverbial sin que participe ningún otro símbolo.
- ◆ La **partícula *se* ante *lo, la, los, las***. Cuando se dan combinaciones del tipo “*se lo dije*”, se reconocería de manera errónea como objetos directos tanto a *se* como a *lo*; sin embargo, *se* ante *lo, la, los, las* no será nunca objeto directo.
- ◆ Subordinadas **introducidas por *de que***. Cuando se tiene una oración subordinada introducida por el conector *de que*, la función de la subordinada nunca será la de circunstancial. Ejemplo “*La idea de que somos amigos me alegra*”.
- ◆ El **pronombre de relativo como conector subordinante**. El pronombre de

relativo puede dar lugar a un *núcleo nominal* o a un *conector subordinante*. Como el *conector subordinante* tiene como función la unión de una proposición subordinada, necesita que la oración contenga al menos dos *núcleos verbales* —uno por proposición.

- ◆ **Acompañantes necesarios**. El *adyacente* debe acompañar a una palabra que pueda ser *núcleo nominal* o a otro modificador del *núcleo nominal*. El *adjetivo* como *núcleo nominal* debe ir acompañado de un presentador que sustantive su función.

- ◆ **La aposición**. Se rechaza la generación de *aposición* —se considera *adyacente*— cuando se genera a partir de un *sustantivo* que puede tener función de *adjetivo*.

- ◆ **El verbo haber**. Cuando aparecen casos como “*hemos salido*” en el que *hemos* puede considerarse el *núcleo verbal* y *salido* un *sustantivo*, se debe tener en cuenta que el verbo *haber* como *núcleo* siempre debe ir acompañado de un objeto directo.

- ◆ **El núcleo verbal pasivo**. El núcleo de una oración pasiva está siempre compuesto por una *forma verbal compleja*, en la que el *auxiliar* es siempre el verbo *ser*.

- ◆ **El núcleo verbal copulativo**. Sólo puede generarse a partir de un conjunto reducido de verbos —*estar, parecer, permanecer, resultar, hacerse, seguir,...*

3 Resolución de otros conflictos.

Además de las reglas anteriores, existe un conjunto de reglas que, sin aplicarse directamente a ningún conflicto primario, sirven para eliminar ambigüedades.

3.1 Símbolos que no pueden cubrir toda la oración.

Toda oración cumple una serie de características en cuanto a lo que a su estructura se refiere: debe tener *sujeto* y *predicado*, y en ciertos casos, tal vez sólo *predicado*. El símbolo *sujeto* no puede cubrir por sí sólo toda la secuencia objeto del análisis.

La *proposición* y la *proposición subordinada* son símbolos que se han definido para generar análisis de oraciones compuestas, es por ello que un símbolo *proposición* teniendo la misma estructura que los símbolos *oración*, no debe nunca cubrir toda la frase.

Regla: Símbolos totales.

El símbolo Δ no terminal que cubre toda la secuencia a analizar es aceptado si y sólo si se encuentra entre los símbolos del conjunto Ω que tienen permitido cubrir toda una secuencia.

3.2 La forma verbal compleja.

Una estructura de perífrasis verbal —formada por más de dos elementos—, genera una *forma verbal compleja* sólo en los siguientes casos, [GOM02]:

<i>acabar de + infinitivo</i>
<i>deber de + infinitivo</i>
<i>dejar de + infinitivo</i>
<i>echarse a + infinitivo</i>
<i>empezar a + infinitivo</i>
<i>estar para + infinitivo</i>
<i>explotar a + infinitivo</i>
<i>haber de + infinitivo</i>
<i>haber que + infinitivo</i>
<i>ir a + infinitivo</i>
<i>llagar a + infinitivo</i>
<i>ponerse a + infinitivo</i>
<i>romper a + infinitivo</i>
<i>tener que + infinitivo</i>
<i>venir a + infinitivo</i>
<i>volver a + infinitivo</i>

Tabla 17: Perífrasis verbales consideradas

3.3 Otros casos.

Existen determinadas situaciones en las que pueden resolverse ambigüedades a partir de consideraciones sobre las palabras, categorías gramaticales y objetos que intervienen:

- ❖ Cuando aparece el adverbio *muy*, no puede dar lugar por sí sólo a ningún tipo de sintagma, pues su función es la de modificar el significado de otras palabras.

- ❖ Las palabras *a* y *de* tienen como posibles funciones la de *preposición* y *sustantivo*. Cuando se encuentran ante un *presentador* su función es la de *preposición*, y no pueden dar lugar a ningún nuevo símbolo por sí solas.

- ❖ La palabra *se* puede aparecer ante el verbo sin una función concreta dentro de la oración; en tal caso, el *núcleo verbal* se encuentra siempre en tercera persona.

- ❖ En oraciones como “*el pájaro que canta muestra su alegría*”, *que* puede funcionar como adyacente; sin embargo, nunca lo hará pospuesto al núcleo nominal, y en tal caso no puede complementar a *pájaro*. Lo mismo ocurre con la palabra *no* en la oración “*como Isabel no nos llame, no la esperaremos*”, donde *no* en ningún caso complementa a *Isabel*.

- ❖ Existe un conjunto de palabras que no podrán dar lugar a un sintagma por sí solas.

Así, las palabras *tan* y *muy* no pueden generar un *sintagma adverbial* si no van acompañadas de otros elementos. Lo mismo ocurre con *algún*, *un*, *unos*, *unas* respecto del *sintagma adjetivo*.

- ❖ Cuando existe una ambigüedad funcional del tipo *adjetivo* y *sustantivo*, de modo que la palabra por sí sola en ambos casos dé lugar a un símbolo *atributo*, es posible descartar el símbolo *atributo* generado a partir del sustantivo.

- ❖ Si un adjetivo se encuentra precedido por un presentador —sólo en el caso de tratarse de presentador que no pueda aparecer pospuesto al núcleo nominal, como el artículo determinado— no puede dar lugar a un sintagma adjetivo.

- ❖ En la oración “*te tengo dicho que no vengas tarde a casa*”, es posible eliminar la ambigüedad de la palabra *tarde* descartando la opción verbo cuando se encuentra tras otro verbo.

- ❖ El *predicativo subjetivo* no puede aparecer con verbos copulativos, pues en tal caso su función sería la de atributo.

- ❖ Si aparece antepuesta una preposición a un *sintagma nominal*, el único símbolo que puede generarse a partir del sintagma nominal es el *sintagma preposicional* correspondiente.

- ❖ En oraciones como “*Juan, que es diligente, trabaja sin descanso*”, se podría reconocer la combinación *sustantivo + conector + sustantivo* como la unión de dos *sintagma nominales homogéneos* para dar lugar a un *sintagma nominal heterogéneo*; sin embargo, para que el conector que une dos sintagmas nominales sea la *coma*, es necesario que al menos uno de los sintagmas sea heterogéneo.

- ❖ Cuando una proposición subordinada tiene función de adyacente, siempre se encontrará pospuesta al núcleo nominal, nunca antepuesta.

- ❖ Cuando se genera un adyacente a partir de un sintagma adjetivo, el sintagma adjetivo estará formado por al menos dos palabras.

- ❖ Si existe un *vocativo*, aparece separado por una coma.

4 Consideraciones para la generación del símbolo predicado.

La existencia de ambigüedades da lugar a múltiples símbolos que pueden no formar ningún árbol final. Esto produce un aumento importante del tiempo de procesamiento. Uno

de los mayores problemas aparece en la generación de símbolos *predicado*. Las reglas que definen las estructuras del predicado vienen dadas por las combinaciones de elementos que puedan aparecer en el mismo. En una definición formal de tipo estructural sería necesario indicar todas las posibles combinaciones; debido a que la colocación de la mayor parte de los elementos es libre, el número de posibles estructuras del predicado sería enorme.

Si el número de elementos del predicado es n , entonces son posibles $n!$ combinaciones. Sean los elementos posibles *núcleo verbal*, *objeto directo*, *objeto indirecto*, *predicativo objetivo*, *predicativo subjetivo*, *atributo*, *complemento agente*, *suplemento*, *circunstancial de tiempo*, *circunstancial de lugar*, *circunstancial de modo*, *circunstancial de fin*, *circunstancial de causa*, *circunstancial de medio*, *circunstancial de argumento*, *circunstancial de compañía*, *circunstancial de negación*, *circunstancial de afirmación*, *circunstancial de cantidad*, *circunstancial de oposición*, *circunstancial de duda*. Se tienen 21 elementos; si se supone que los dos únicos elementos que pueden repetirse en un predicado simple son el *objeto directo* y el *objeto indirecto* —se ha considerado que los circunstanciales no se repiten para calcular una cota inferior del máximo número de combinaciones posibles—, se obtiene un número de posibles combinaciones según las variaciones de 23 elementos tomados de n en n para n desde 1 hasta 23 y sumándolas.

$$\sum_{i=1}^{23} V_{23,i} = \sum_{i=1}^{23} \frac{23!}{(23-i)!}$$

Sin embargo, todas no se pueden dar por lo que no es necesario añadirlas como reglas; existen combinaciones de diferentes elementos que no pueden aparecer formando parte de un mismo predicado, como se muestra en la tabla 18.

Aparte de las reducciones que puedan suponer las reglas de desambiguación vistas hasta ahora en la no generación de símbolos predicado, es posible aprovechar ciertas características del predicado. En las reglas en las que aparece el predicado se encuentra delimitando —por el principio o el final— el símbolo que genera, o bien es adyacente a un símbolo de puntuación. Es por ello que se deben rechazar todas aquellas combinaciones de símbolos, que tengan como finalidad la

posible generación de símbolos *predicado*, tal que alguno de los extremos no es principio o final del símbolo generado o exista un símbolo de puntuación adyacente.

atributo	atributo
atributo	morfema de pasiva
atributo	objeto directo
atributo	objeto indirecto
atributo	predicativo objetivo
atributo	predicativo subjetivo
atributo	suplemento
complemento agente	suplemento
morfema de impersonal	morfema de impersonal
morfema de impersonal	morfema de pasiva
morfema de impersonal	morfema de voz media
morfema de impersonal	predicativo subjetivo
morfema de pasiva	atributo
morfema de pasiva	morfema de impersonal
morfema de pasiva	morfema de pasiva
morfema de pasiva	morfema de voz media
morfema de pasiva	objeto directo
morfema de pasiva	objeto indirecto
morfema de voz media	morfema de impersonal
morfema de voz media	morfema de pasiva
morfema de voz media	morfema de voz media
núcleo verbal	núcleo verbal
núcleo verbal	núcleo verbal copulativo
núcleo verbal	núcleo verbal pasivo
núcleo verbal copulativo	núcleo verbal
núcleo verbal copulativo	núcleo verbal copulativo
núcleo verbal copulativo	núcleo verbal pasivo
núcleo verbal pasivo	núcleo verbal
núcleo verbal pasivo	núcleo verbal copulativo
núcleo verbal pasivo	núcleo verbal pasivo
objeto directo	atributo
objeto directo	morfema de pasiva
objeto indirecto	atributo
objeto indirecto	morfema de pasiva
predicativo objetivo	atributo
predicativo objetivo	predicativo objetivo
predicativo objetivo	predicativo subjetivo
predicativo subjetivo	atributo
predicativo subjetivo	morfema de impersonal
predicativo subjetivo	predicativo objetivo
predicativo subjetivo	predicativo subjetivo
suplemento	atributo
suplemento	complemento agente
suplemento	suplemento

Tabla 18: Parejas de símbolos incompatibles en un mismo predicado

En ocasiones, esta condición puede no cumplirse. Es el caso de la existencia de oraciones subordinadas, en estos casos antes de rechazar un evento debe tenerse en cuenta la posibilidad de que se trate de una oración subordinada, para ello se verifica la existencia de elementos subordinantes y la existencia de múltiples formas verbales.

Hay un conjunto de símbolos a los cuales es posible aplicarles reglas similares. Así, la proposición ya sea coordinada o subordinada debe llevar antepuesto o pospuesto, según sea el caso, algún tipo de conector, ya sea subordinante o coordinante. De igual manera cualquier tipo de oración que vaya a dar lugar a una de estas proposiciones debe cumplir estas mismas condiciones. Teniendo en cuenta

Combinación de símbolos que puede dar lugar a más de un símbolo																				
presentador + núcleo nominal + adyacente + complemento preposicional del nombre				X																
presentador + núcleo nominal + adyacentes				X																
presentador + núcleo nominal + complemento preposicional del nombre				X																
pronombre personal átono	X			X																
pronombre de relativo		X															X			
otro pronombre		X				X							X							
proposición				X	X	X	X	X												
proposición de infinitivo				X	X	X	X	X												
proposición subordinada				X	X	X	X	X												
puntuación + oración atributiva + puntuación																				X
puntuación + oración de suplemento + puntuación																				X
puntuación + oración intransitiva + puntuación																				X
puntuación + oración pasiva + puntuación																				X
puntuación + oración transitiva + puntuación																				X
puntuación + sentencia + puntuación																				X
sintagma adjetivo				X	X	X														
sintagma nominal heterogéneo				X	X	X	X	X												
sintagma nominal homogéneo				X	X	X	X	X												
sintagma nominal sustitutivo				X	X	X	X	X												
sintagma preposicional	X			X	X	X	X	X												
sujeto + coma + predicado				X																
sujeto + predicado				X																
sujeto + predicado + puntuación				X																
sustantivo																			X	
verbo					X															

Tabla 20b: Resumen de aplicación de reglas por ambigüedad estructural directa

6 Ejemplo de aplicación de las reglas de resolución de conflictos primarios.

Para la ambigüedad estructural directa formada por las reglas de la gramática:

- o atributo → sintagma nominal homogéneo
- o circunstancial → sintagma nominal homogéneo
- o objeto directo → sintagma nominal homogéneo
- o sujeto → sintagma nominal homogéneo
- o vocativo → sintagma nominal homogéneo

Se aplican las siguientes reglas de resolución de conflictos primarios:

Simbolos incompatibles. El *atributo* es incompatible con la existencia de otro *atributo*, con la de un *morfema de pasiva*, con la de un *objeto directo*, con la de un *objeto indirecto*, con la de una *oración intransitiva*, con la de una *oración transitiva*, con la de un *predicativo objetivo*, con la de un *predicativo subjetivo* y con la de un *suplemento*. El *objeto directo* es incompatible con la existencia de un *atributo*, con la de un *morfema de pasiva*, con la de una *oración de suplemento* y con la de una *oración intransitiva*. El *sujeto* es incompatible con la existencia de un *morfema de impersonal*.

Símbolos necesarios. El *atributo* requiere la existencia de un *núcleo verbal copulativo* y el *objeto directo* la de un *núcleo verbal*.

Concordancias. El *sujeto* debe concordar en número y persona con el *núcleo verbal*.

Semántica necesaria. Cabe desechar ciertos circunstanciales —lugar, tiempo y cantidad— en función de la información semántica de sus palabras.

Casos especiales: Doble objeto directo. De existir otro *objeto directo*, se descarta la generación de *objeto directo* a menos que el primero le siga en la oración, esté generado a partir de un *pronombre personal átono* y sus núcleos concuerden en género y número.

7 Ambigüedades reales.

Se considera la aparición de ambigüedades reales en aquellos casos en los que se obtiene como resultado múltiples análisis sintácticos. Todos deben ser aceptados, ya que algunos autores interpretan de diferente manera la función de alguna de sus partes.

Algunos ejemplos serían:

“Hablaron los representantes de los municipios”.

El complemento “de los municipios” puede estar modificando al núcleo del *sujeto*, “representantes”, o tener función de *suplemento* como parte del *predicado verbal simple*.

hablaron	los	representantes	de	los	municipios
<i>verbo</i>	<i>artículo</i>	<i>sustantivo</i>	<i>preposición</i>	<i>artículo</i>	<i>sustantivo</i>
<i>forma verbal simple</i>	<i>presentador</i>	<i>núcleo nominal</i>	<i>preposición subordinante</i>	<i>presentador</i>	<i>núcleo nominal</i>
<i>núcleo verbal</i>	<i>sintagma nominal homogéneo</i>				
	<i>sintagma preposicional</i>				
	<i>complemento preposicional del nombre</i>				
	<i>sintagma nominal heterogéneo</i>				
<i>predicado</i>	<i>sujeto</i>				
<i>oración intransitiva</i>					

Figura 12: Interpretación A

hablaron	los	representantes	de	los	municipios
<i>verbo</i>	<i>artículo</i>	<i>sustantivo</i>	<i>preposición</i>	<i>artículo</i>	<i>sustantivo</i>
<i>forma verbal simple</i>	<i>presentador</i>	<i>núcleo nominal</i>	<i>preposición subordinante</i>	<i>presentador</i>	<i>núcleo nominal</i>
<i>núcleo verbal</i>	<i>sintagma nominal homogéneo</i>			<i>sintagma nominal homogéneo</i>	
	<i>sintagma preposicional</i>				
	<i>suplemento</i>				
<i>predicado</i>	<i>sujeto</i>		<i>predicado</i>		
<i>oración intransitiva</i>					

Figura 13: Interpretación B

“Llenaron el vaso de agua”.

El complemento “de agua” puede estar modificando al núcleo del *objeto directo*, “vaso”, o tener función de *suplemento* como parte del *predicado verbal simple*.

llenaron	el	vaso	de	agua
<i>verbo</i>	<i>artículo</i>	<i>sustantivo</i>	<i>preposición</i>	<i>sustantivo</i>
<i>forma verbal simple</i>	<i>presentador</i>	<i>núcleo nominal</i>	<i>preposición subordinante</i>	<i>núcleo nominal</i>
<i>núcleo verbal</i>				<i>sintagma nominal homogéneo</i>
				<i>sintagma preposicional</i>
				<i>complemento preposicional del nombre</i>
				<i>sintagma nominal heterogéneo</i>
				<i>objeto directo</i>
				<i>predicado</i>
				<i>oración intransitiva</i>

Figura 14: Interpretación A

llenaron	el	vaso	de	agua
<i>verbo</i>	<i>artículo</i>	<i>sustantivo</i>	<i>preposición</i>	<i>sustantivo</i>
<i>forma verbal simple</i>	<i>presentador</i>	<i>núcleo nominal</i>	<i>preposición subordinante</i>	<i>núcleo nominal</i>
<i>núcleo verbal</i>	<i>sintagma nominal homogéneo</i>			<i>sintagma nominal homogéneo</i>
				<i>sintagma preposicional</i>
				<i>objeto directo</i>
				<i>suplemento</i>
				<i>predicado</i>
				<i>oración intransitiva</i>

Figura 15: Interpretación B

“Esa ruta cubre trescientos kilómetros”

La parte de la oración “trescientos kilómetros” puede ser reconocida como *circunstancial de distancia, circunstancial de cantidad u objeto directo* —sujeto cuando se convierte en pasiva.

esa	ruta	cubre	trescientos	kilómetros
<i>demonstrativo</i>	<i>sustantivo</i>	<i>verbo</i>	<i>adjetivo</i>	<i>sustantivo</i>
<i>presentador</i>	<i>núcleo nominal</i>	<i>forma verbal simple</i>	<i>adyacente</i>	<i>núcleo nominal</i>
<i>sintagma nominal homogéneo</i>		<i>núcleo verbal</i>	<i>sintagma nominal homogéneo</i>	
			<i>circunstancial de distancia</i>	
<i>sujeto</i>			<i>predicado</i>	

oración intransitiva

Figura 16: Interpretación A

esa	ruta	cubre	trescientos	kilómetros
<i>demonstrativo</i>	<i>sustantivo</i>	<i>verbo</i>	<i>adjetivo</i>	<i>sustantivo</i>
<i>presentador</i>	<i>núcleo nominal</i>	<i>forma verbal simple</i>	<i>adyacente</i>	<i>núcleo nominal</i>
<i>sintagma nominal homogéneo</i>		<i>núcleo verbal</i>	<i>sintagma nominal homogéneo</i>	
			<i>circunstancial de cantidad</i>	
<i>sujeto</i>			<i>predicado</i>	

oración intransitiva

Figura 17: Interpretación B

esa	ruta	cubre	trescientos	kilómetros
<i>demonstrativo</i>	<i>sustantivo</i>	<i>verbo</i>	<i>adjetivo</i>	<i>sustantivo</i>
<i>presentador</i>	<i>núcleo nominal</i>	<i>forma verbal simple</i>	<i>adyacente</i>	<i>núcleo nominal</i>
<i>sintagma nominal homogéneo</i>		<i>núcleo verbal</i>	<i>sintagma nominal homogéneo</i>	
			<i>objeto directo</i>	
<i>sujeto</i>			<i>predicado</i>	

oración transitiva

Figura 18: Interpretación C

8 Resultados estadísticos.

Se ha realizado el análisis de 776 oraciones seleccionadas que cubrieran el mayor espectro de la casuística inherente a la gramática española. Se presentan como estadísticas el número de nodos —símbolos— generados durante el proceso de generación de árboles de representación sintáctica con la desambiguación funcional local —basada en estructuras de carácter local— y sin ella. Ante el coste computacional del proceso sin desambiguación funcional se han analizado las oraciones de hasta 9 símbolos para la comparación.

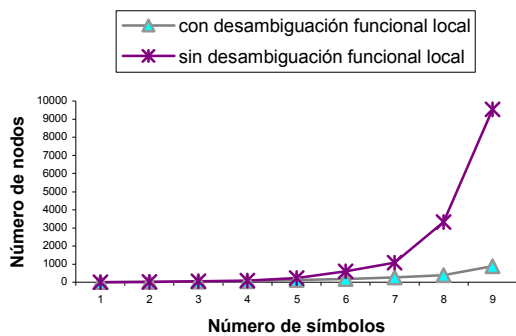


Figura 19: Nodos generados durante el proceso de generación de árboles de representación sintáctica

Puede verse que el número de nodos que se generan durante el proceso de análisis —medida del coste de los procesos de generación de árboles— disminuye al aplicar la desambiguación funcional local y, en general, aumenta con el número de símbolos de la oración.

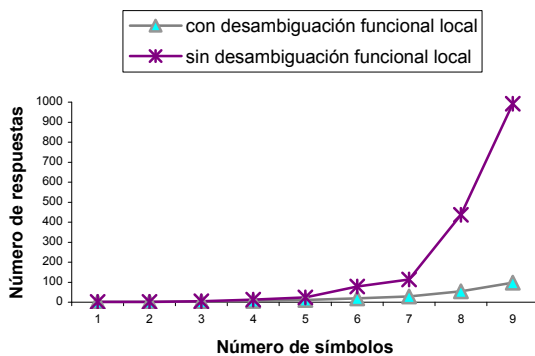


Figura 20: Resultados

El número de respuestas —número de árboles de análisis diferentes— disminuye al aplicar la *desambiguación funcional local*,

porque la *desambiguación funcional local* tiene en cuenta aspectos que no se tratan en la *desambiguación estructural* —las *combinaciones vedadas* y el conjunto de restricciones por frecuencia— que eliminan gran cantidad de posibilidades.

Estudiada la influencia de la etapa de la desambiguación funcional local, se pasa a estudiar el comportamiento del generador de árboles de representación sintáctica con la desambiguación funcional local; de esta manera, se analizan frases hasta un número de símbolos mayor. En las siguientes gráficas se observa cómo el número de nodos y el de respuestas tienden a aumentar de manera gradual con el número de símbolos; sin embargo, aparecen picos y valles que responden a las características de las oraciones —aumentando sobre todo cuando se trata de oraciones compuestas— y no sólo al número de símbolos de la oración y al número de combinaciones funcionales analizadas.

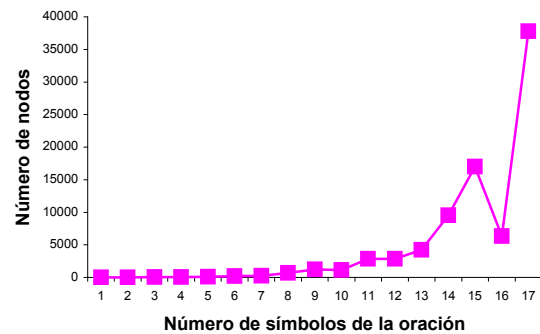


Figura 21: Número de nodos generados durante el análisis de texto real

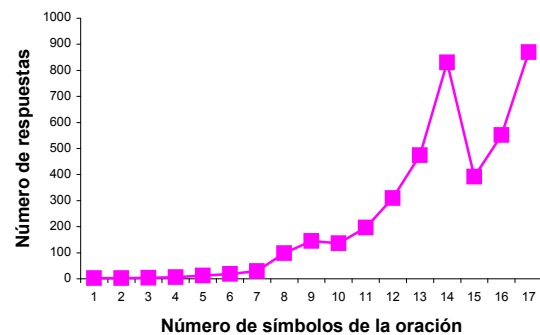


Figura 22: Número de respuestas generadas durante el análisis de texto real

Otro dato de interés radica en la influencia de la generación de árboles de representación sintáctica en la desambiguación funcional. Una vez que se realiza el proceso de análisis, el número de combinaciones aceptadas se reduce; la bondad en el proceso de desambiguación funcional del 87% crece hasta un 96%. En la siguiente gráfica puede observarse cómo afecta a los resultados la aplicación de los procesos de desambiguación estructural sobre la funcional.

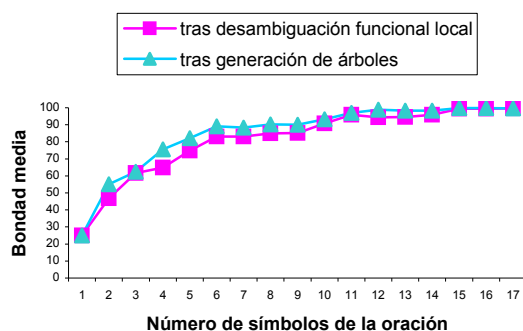


Figura 23: Influencia de la desambiguación estructural en la bondad de la desambiguación funcional

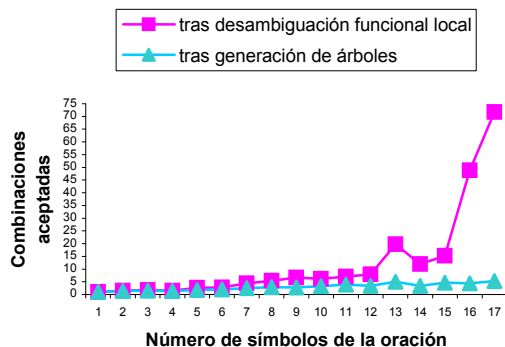


Figura 24: Influencia de la desambiguación estructural en el número de combinaciones aceptadas de la desambiguación funcional

9 Conclusiones.

Este trabajo trata la gramática española desde un punto de vista estructural con reglas simples que luego se unen a un conjunto de métodos de desambiguación; no se detiene en subconjuntos de la gramática como la mayor parte de los trabajos, sino que apuesta por un sistema de reglas con las posibles combinaciones de la gramática española a pesar de la notable cantidad de combinaciones necesarias.

Se identifica y se aporta solución al problema de la aparición de ambigüedades

estructurales que se generan durante el proceso de construcción de árboles de representación sintáctica. Las estructuras sintácticas se combinan entre sí para dar lugar a los árboles de representación sintáctica —muchas de estas combinaciones generan árboles erróneos. Se han identificado los conflictos directos entre reglas como una de las causas principales del problema y se han desarrollado métodos de desambiguación estructural, para ello se han estudiado las características de las diferentes estructuras sintácticas y cómo deben tenerse en cuenta a la hora de aceptar o no la construcción de un símbolo de representación.

A la vista del gran número de combinaciones posibles de los elementos de la gramática —se hace más patente en las construcciones del predicado, por permitirse cualquier número de elementos y casi cualquier combinación—, se han definido los mecanismos adecuados de representación de modo que queden cubiertas todas las posibilidades, sin dejar de representar a opciones válidas. Al permitir cualquier combinación de elementos posibles en el predicado aparecen combinaciones que no deberían ser permitidas y que serían rechazadas en los procesos de desambiguación estructural. Quedan así representadas todas las combinaciones posibles desde un punto de vista estructural y rechaza las no permitidas.

Entre las características estudiadas se consideran las relaciones entre las diferentes estructuras, las características semánticas de éstas, las características de grupos de palabras asociadas a algunas categorías funcionales —como las preposiciones—, y las de algunas palabras concretas.

Se catalogan grupos de identificación semántica orientados al reconocimiento de estructuras sintácticas. Entre los procesos de desambiguación estructural se encuentran algunas reglas que introducen información semántica. Las listas generadas se han obtenido a partir de los cuadros de los diccionarios ideológicos que pueden relacionarse con ciertas estructuras sintácticas.

10 Bibliografía.

[BIA98] Bianchi De Cortina, E. 1998. Teoría Gramatical (Gramática Estructural). Ediciones Daly S.L. Córdoba.

- [BOS99] Bosque, I., Demonte, V. 1999. Gramática descriptiva de la lengua española. 3 vols. Espasa. Madrid.
- [GIL98] Gili Gaya, S. 1998. Curso Superior de Sintaxis Española. Biblograf, s.a. Barcelona.
- [GOM96] Gómez Guinovart, J. 1996. Aportaciones a la metodología de evaluación de los sistemas de verificación automática de la sintaxis. Universidad de Vigo.
- [GOM02] Gómez Torrego, L. 2002. Gramática didáctica del español. S.M. Madrid.
- [GOM02] Gómez Torrego, L. 2002. Análisis sintáctico. Teoría y práctica. S.M. Madrid.
- [GUT78] Gutiérrez Araus, M. L. 1978. Estructuras sintácticas del español actual. Sociedad General Española de Librería, S.A. Madrid.
- [ONV90] Onieva Morales, J. L. 1990. Cómo dominar el análisis gramatical superior. Playor. Madrid.
- [QUE96] Quesada, J. F. 1996. Un modelo robusto y eficiente para el análisis sintáctico de lenguajes naturales mediante árboles múltiples virtuales. Centro Informático Científico de Andalucía (CICA).
- [QUE97] Quesada, J. F. 1997. El Algoritmo SCP de Análisis Sintáctico mediante Propagación de Restricciones. Tesis Doctoral. Universidad de Sevilla. Junio de 1997.
- [RAE89] Real Academia Española. 1989. Esbozo de una nueva gramática de la lengua española. Espasa-Calpe. Madrid.
- [ROD96] Rodríguez Magro, C. 1996. CORRECTOR: Un sistema de verificación sintáctica y estilística de textos. Centro de Investigación UAM-IBM (Madrid), Centro de Tecnología de la Lengua (Sevilla).
- [ROD92] Rodríguez, C.; Rubio, C.; Sánchez, A.; Sopeña, L. 1992. Herramientas de ayuda a la redacción de textos: Un sistema de verificación léxica, sintáctica y estilística. Voz y Letra, Tomo III, Volumen 1. Centro de Tecnología de la lengua-IBM, Sevilla.
- [ROD97] Rodríguez Magro, C.; Sopeña, L. Villar, C. 1997. Caracterización sintáctica de los verbos castellanos para una gramática en ordenador. Procesamiento del Lenguaje Natural. Boletín nº 7.
- [SAN02] Santana Suárez, O.; Pérez Aguiar, J.; Losada García, L.; Carreras Riudavets, F. 2002. Hacia la desambiguación funcional automática en Español. Procesamiento del Lenguaje Natural, nº 28, SEPLN.
- [SAN97] Santana, O., J. Pérez, Z. Hernández, F. Carreras y G. Rodríguez. 1997. FLAVER: Flexionador y lematizador automático de formas verbales. Lingüística Española Actual XIX, 2: 229-282.
- [SAN99] Santana, O.; J. Pérez, F. Carreras, J. Duque, Z. Hernández y G. Rodríguez. 1999. FLANOM: Flexionador y lematizador automático de formas nominales. Lingüística Española Actual XXI, 2: 253-297.
- [ZAE95] Zaenen, A.; Uszkoreit, H.; Karlsson, F.; Karttunen, L.; Sanfilippo, A.; Pulman, S. G.; Pereira, F.; Briscoe, T. 1995. Language Analysis and Understanding. Survey of the Art in Human Language Technology. Capítulo 3.