

Restricciones sobre orden de constituyentes basadas en corpus y destinadas a la desambiguación sintáctica*

M. Pilar Valverde Ibáñez
Departamento de Lingua Española
Universidade de Santiago de Compostela
Av. Burgo das Nacións, s/n
15782 Santiago de Compostela
mpilarv@usc.es

Resumen: El español es una lengua de orden de constituyentes oracionales considerablemente flexible. Sin embargo, el orden en que aparecen las funciones no es arbitrario sino que está regido por factores de distinto tipo, algunos de ellos detectables formalmente. En este artículo se presenta una propuesta de descripción del orden de los constituyentes oracionales en español basada en corpus y destinada a la desambiguación de árboles de representación sintáctica.

Palabras clave: orden de constituyentes, desambiguación, ambigüedad estructural, análisis sintáctico

Abstract: Spanish is a quite flexible constituent order language. However, the order in which functions appear is not arbitrary but is due to several factors, some of them formally detectable. This paper presents a proposal of description of the order of clause-level constituents in Spanish, based in corpus and aimed at the desambiguation of syntactic trees.

Keywords: constituent order, disambiguation, structural ambiguity, syntactic analysis

1. Introducción

La ambigüedad, es decir, la posibilidad de interpretar una secuencia de más de una forma, es un problema fundamental en el procesamiento del lenguaje natural, y más concretamente en el análisis sintáctico automático. En este ámbito, nos enfrentamos a dos tipos de ambigüedad: la ambigüedad real, propia del lenguaje natural, y la ambigüedad resultante del análisis automático, que se produce por un sobregeneración de análisis.

En el análisis sintáctico del español, por ejemplo, si partimos de un análisis en constituyentes correctamente segmentados y etiquetados con su categoría sintáctica, y les asignamos las funciones que potencialmente pueden desempeñar, el resultado será prácticamente siempre, una secuencia ambigua, lo que supone un problema para los analizadores sintácticos, como confirman Gala (1999), Pineda y Meza (2003) y Santana et al. (2004).

Por ejemplo, si tomamos una secuencia de la forma <pronombre demostrativo + grupo

verbal + sintagma preposicional introducido por *a* y seguido por un determinante y un nombre común> y asignamos a cada constituyente las funciones oracionales que potencialmente puede desempeñar (sin tener en cuenta la función de complemento circunstancial, por los motivos que explicamos en el apartado 2), el resultado es ambiguo, ya que por ejemplo un pronombre demostrativo puede funcionar como sujeto, complemento directo o complemento predicativo del sujeto. En el cuadro 1 mostramos la correspondencia entre categorías y funciones sintácticas que se da en el corpus que utilizamos como referencia.¹ Cada fila representa una categoría y cada columna, una función. El significado de las siglas se muestra en el anexo A.

En el caso de los sintagmas preposicionales, hay que tener en cuenta que el complemento directo y el indirecto sólo pueden estar introducidos por la preposición *a*, que los complementos predicativos suelen estar introducidos por *como*, y el complemento agente, por *por*. El complemento de régimen, en cambio, puede estar introducido por un conjunto más amplio de preposiciones. Podemos ver

* Este trabajo ha sido realizado gracias a la beca FPI BES-2004-4725 asociada al proyecto BFF2003-02866 del Ministerio de Educación y Ciencia. También quiero agradecer a Montserrat Cívit Torruella sus comentarios sobre la versión preliminar del artículo.

¹La Base de Datos Sintácticos del español actual, de la que damos más detalles en el apartado 3.

Categorías	S	D	I	R	P	O	A	i
sn:det.nc	x	x			x	x		4
sn:det.np	x	x			x	x		4
sn:det.x	x	x			x	x		4
sn:np	x				x	x		3
sn:nc	x	x			x	x		4
sn:pp	x				x			2
sn:pin	x	x			x	x		4
sn:pd	x	x			x			3
sn:px					x	x		2
S.F.C.que	x	x			x	x		4
S.F.C.Q	x	x			x	x		4
S.F.C.si	x	x						2
S.F.A	x	x		x	x	x		5
S.NF.C	x	x			x	x		4
S.NF.A					x	x		2
S.NF.Q	x	x						2
S.NF.si		x						1
sn:det.sfr	x	x			x			3
S.F.R-	x	x			x			3
sadj					x	x		2
S.NF.P					x	x		2
sadv		x		x	x	x		4
p+sn:det.nc		x	x	x	x	x	x	6
p+sn:det.np		x	x	x	x	x	x	6
p+sn:det.x		x	x	x	x	x	x	6
p+sn:np		x	x	x	x	x	x	6
p+sn:nc		x	x	x	x	x	x	6
p+sn:pp		x	x	x	x	x	x	6
p+sn:pin		x	x	x	x	x	x	6
p+sn:pd		x	x	x	x	x	x	6
p+sn:px						x		1
p+sn:det.sfr		x	x	x	x	x	x	6
p+S.F.R		x	x	x	x	x	x	6
p+S.NF.C			x	x				2
p+S.F.C.que				x				1
p+S.F.C.Q				x				1
p+S.F.C.si				x				1
p+S.F.A				x		x		2
p+sadv				x	x			2
p+sadj				x	x	x		3
p+S.NF.P				x	x	x		3

Cuadro 1: Categorías sintácticas y las funciones argumentales que pueden desempeñar

que prácticamente todas las categorías son ambiguas en cuanto a su función sintáctica, ya que tienen más de una interpretación posible.² De media, a cada categoría le corresponden 3,56 interpretaciones distintas.

Si retomamos la secuencia vista anteriormente, de la forma <pronombre demostrativo + grupo verbal + sintagma preposicional introducido por *a* y seguido por un determinante y un nombre común>, teniendo en cuenta que ninguna función argumental puede repetirse, dicha secuencia recibiría 8 análisis distintos (SVD, SVI, SVR, DVI, DVR,

²En el cuadro, la columna “i” indica el número de interpretaciones distintas para cada categoría.

PVD, PVI, PVR), ya que el pronombre demostrativo puede ser S, D o P, el sintagma preposicional introducido por *a* y seguido por un determinante y un nombre común puede ser D, I o R, y al grupo verbal siempre le asignamos la función “predicado” (V).

2. Objeto de estudio

El conocimiento lingüístico formalmente identificable y destinado a la desambiguación sintáctica trata básicamente sobre la concordancia, las preposiciones y el orden de los constituyentes.³ Este último aspecto es el que tratamos en este artículo. Concretamente, delimitamos nuestro objeto de estudio a:

- El orden de las funciones dentro de una misma cláusula. No tratamos los casos en que un constituyente se desplaza fuera de la cláusula a la que pertenece, como en el ejemplo 1, en que el complemento de régimen *a la iglesia* se encuentra fuera de la cláusula de infinitivo a la que pertenece.⁴

(1) *A la iglesia*, por supuesto, estaba totalmente prohibido *entrar sin medias o con manga corta*.

- El orden de los argumentos y el predicado. Si dejamos a un lado los elementos textuales y modificadores oracionales del tipo *en cuanto a, francamente, etc.*, podemos diferenciar dos tipos de funciones oracionales:

- Los argumentos y el predicado: forman el esquema verbal y sólo pueden aparecer una vez. El hecho de que el complemento directo y el indirecto puedan coaparecer con un clítico acusativo y dativo, respectivamente, no invalida esta afirmación, ya que no consideramos los clíticos como funciones explícitas, con el mismo estatus que el resto de constituyentes, sino como parte del grupo verbal.

³No tratamos aquí el conocimiento semántico o pragmático.

⁴Para más detalles sobre este fenómeno, ver por ejemplo Torrego (1984), que analiza la extraposición de sintagmas relativos y sus limitaciones en español y Holan et al. (2000), que analizan este fenómeno y sus distintas repercusiones en el análisis sintáctico automático del inglés y el checo.

- Los complementos circunstanciales: pueden aparecer en distinto número. Además, según diversos autores, como Gutiérrez (1978) y López García (1996), presentan un orden más libre que las funciones argumentales.

Para homogeneizar el campo de estudio, por el momento estudiamos solamente la posición de los argumentos y el predicado. Las funciones tradicionalmente consideradas como argumentos en español son: sujeto (S), complemento directo (D), indirecto (I), de régimen (R), predicativo del sujeto (P), predicativo del complemento directo (O) y agente (A). El predicado (V) es la función que asignamos al grupo verbal.

- Las funciones explícitas, y no en forma de clítico, ya que los clíticos siempre aparecen inmediatamente delante o detrás del verbo. No estudiamos su posición, pero sí que nos interesa su presencia en la secuencia, ya que puede indicarnos la posición o presencia de otro complemento, como veremos en el apartado 4.3.
- Las cláusulas que contienen predicado, ya que consideramos este elemento como el eje alrededor del cual se colocan las demás funciones. Por lo tanto, no han sido tenidas en cuenta las cláusulas que no contienen predicado, tanto si se debe a la coordinación de cláusulas paralelas (*gapping*) como al tipo de texto (titulares, etc.).

3. Método

Dado que nuestro objetivo es describir el orden de las funciones argumentales en español desde un punto de vista formal, nos basamos en corpus anotados sintácticamente, y cuando es posible, contrastamos las conclusiones extraídas con la bibliografía existente (Fernández Ramírez (1951), Fish (1960), Gili Gaya (1961), Butt y Benjamin (1988), Delbecque (1991), por ejemplo). Sin embargo, hay que señalar que existen pocos estudios sobre orden de constituyentes basados en corpus, y que la mayoría trata sobre la posición del sujeto, debido a la considerable movilidad de esta función.

El corpus utilizado ha sido la Base de

Datos Sintácticos del español actual (BDS)⁵ (Rojo, 2001). Se trata de una base de datos que contiene distinta información sintáctica, repartida en 63 campos, sobre 160.000 cláusulas, cada una de las cuales constituye un registro de la base de datos. Las cláusulas analizadas son un subconjunto de un millón y medio de palabras del corpus ARTHUS de la Universidad de Santiago de Compostela.

Para la descripción del orden de los constituyentes oracionales creemos que el método de descripción más adecuado consiste, primero, en establecer grupos de cláusulas que se comportan de forma semejante respecto al orden y, en segundo lugar, describir el orden de constituyentes que presenta cada grupo. Para simplificar, no vamos a entrar en la clasificación de cláusulas.⁶ Para la descripción del orden de las funciones de cada grupo de cláusulas, vamos a proceder de la siguiente manera:

1. Extracción, a partir de los *órdenes documentados* (OD) en el corpus, de restricciones en cuanto al número y tipo de argumentos que aparecen alrededor del predicado.
2. Cálculo de los *órdenes matemáticamente posibles* (OM).
3. Aplicación de las restricciones sobre los órdenes matemáticamente posibles. La eliminación de los órdenes que no cumplen las restricciones da lugar a los órdenes que denominamos *lingüísticamente posibles* (OL). Por lo tanto, el punto de partida son todos los órdenes posibles y la función de las restricciones es eliminar los análisis incorrectos.

$$\text{OM} + \text{restricciones extraídas de OD} = \text{OL}$$

4. Extracción de restricciones adicionales: categoría de las funciones desplazadas, relación entre la presencia de clíticos en el predicado y la posición de ciertas funciones, etc.

No obstante, hemos de tener en cuenta que las lenguas son sistemas abiertos y no existe

⁵Parcialmente accesible en <http://www.bds.usc.es/>

⁶Para más detalles, puede consultarse Valverde, Santalla, y Civit (2005).

una frontera clara entre lo gramatical y lo agramatical. Por lo tanto, la mayoría de restricciones no deben ser entendidas como reglas universales, de gramaticalidad, que siempre se cumplen, sino como tendencias, que *prácticamente* siempre se cumplen en el corpus objeto de estudio, y por lo tanto, como hipótesis, que pueden ser testadas a medida que dispongamos de nuevos corpus anotados como, Cast3LB (Civit y Martí, 2002). De forma preliminar hemos establecido un margen de error máximo del 1 %. Las restricciones no tienen un orden de aplicación concreto sino que son independientes, aunque ciertos órdenes puedan ser rechazados por más de una restricción a la vez.

4. Resultados

Ejemplificamos la aplicación del método anterior a las cláusulas más frecuentes en español: las cláusulas activas personales finitas (exceptuando las cláusulas imperativas) y opcionalmente introducidas por una conjunción subordinante,⁷ como las de los ejemplos 2-5.

- (2) *El guitarrista abandonó el grupo que le hizo famoso en el año 1977.*
- (3) *Sospecho que vas a intentar convencer a esa criatura de que se exhiba en la manifestación.*
- (4) *No sé si me pierdo gran cosa.*
- (5) *El ómnibus estaba donde lo dejamos y el chófer conversaba animadamente con un policía de uniforme verde.*

Pertenecen a este grupo las cláusulas con un verbo finito, en modo indicativo o subjuntivo, personal y activo, es decir, que no contienen la perífrasis <ser + participio> ni el clítico pasivo *se*. Además, este tipo de cláusulas pueden estar introducidas opcionalmente por una conjunción subordinante.

4.1. Restricciones

A partir de los datos de la BDS sobre este tipo de cláusulas, hemos extraído las siguientes restricciones sobre el número y tipo de argumentos que aparecen alrededor del predicado:

- 0.1 No aparece A, ya que es incompatible con un predicado en forma personal. Por

⁷En la BDS estas cláusulas representan aproximadamente un 60 % del total.

lo tanto, las únicas funciones que aparecen son: S, V, D, I, R, P, O.

- 0.2 Contienen como mucho cuatro funciones.
- 0.3 Ninguna función se repite.
- 0.4 La cláusula contiene un predicado.
- 0.5 No aparecen P y O en la misma cláusula, ya que estas dos funciones son incompatibles.

1. Delante del predicado aparece como mucho una función.

Lo más habitual en la BDS es que delante del predicado no aparezca ninguna función (en el 65,19 % de casos), o sólo una función (34,54 %). Esa función es el sujeto en la mayoría de casos (91,28 %) (6), pero también pueden aparecer delante del predicado el resto de funciones: complemento directo (7), complemento indirecto (8), complemento de régimen (9), complemento predicativo del sujeto (10) y complemento predicativo del complemento directo (11).

- (6) *El guitarrista abandonó el grupo que le hizo famoso en el año 1977.*
- (7) *De repente, a Renato le alarma en el viejo cierta expresión doliente.*
- (8) *A él también le aconsejaron no dormir en su casa.*
- (9) *A la reunión, que se celebró en el País Vasco, acudió en representación de la Federación Gallega Pedro Alfaro, secretario general de dicha entidad.*
- (10) *¿Tan descontenta estás de ti misma?*
- (11) *Así los encontró Purísima Vicario a las tres de la madrugada del martes, cuando el alcalde la llevó a despedirse de ellos.*

2. Detrás del predicado aparecen como mucho dos funciones.

En la BDS, lo más frecuente es que detrás del predicado aparezca una función (66,44 %), ninguna (26,29 %) o dos (7,18 %).

3. Si detrás del predicado aparecen dos funciones distintas del sujeto, al menos una

de ellas es un complemento directo o un complemento de régimen, como en el ejemplo 12.

- (12) El proyecto holandés para la unión monetaria brinda [escasas concesiones] [a las economías menos desarrolladas] (SVDI)

El cuadro 2 muestra la precisión de las restricciones anteriores en la BDS. Por ejemplo, para la restricción 0.2 encontramos sólo 13 excepciones (en 9 órdenes distintos), en que la cláusula no está formada por como mucho cuatro elementos sino por cinco, como en el ejemplo 13.

- (13) [A mí] [este] [me lo traje] [de Cuba] [un marinero, que es almirante y es embajador, y cada mañana desde La Habana me manda un barco con una flor] (IDVRS)

Para la restricción 1 encontramos 273 excepciones (en 29 órdenes distintos), en que delante del predicado aparecen no una función sino dos, como en el ejemplo 14.

- (14) [A mí] [ese tipo de show] me gusta mucho. (ISV)

En la restricción 2 encontramos 87 excepciones (en 41 órdenes), en que detrás del predicado aparecen no dos sino tres argumentos, como en el ejemplo 15.

- (15) Salgo [muy mal] [yo] [en las fotos]. (VPSR)

Por último, en la restricción 3 encontramos 5 excepciones (en 2 órdenes distintos), en que detrás del predicado aparecen dos argumentos distintos del sujeto pero ninguno de ellos es complemento directo o de régimen, como en el ejemplo 16.

- (16) El exceso le cae [mal] [a Rusca, que protesta mordisqueando en carne viva]. (SVPI)

Podemos ver que ninguna restricción supera el margen de error de un 1% que hemos establecido, y que la media de error es de un 0,37%.⁸ Por otra parte, todas las restricciones excepto la 3 se aplican a todas las cláusulas que tratamos, que son 99204. La restricción 3 se aplica sólo a aquellas cláusulas que contienen dos elementos distintos del sujeto detrás del predicado.

⁸En el cuadro, el total de excepciones distintas es igual a 365, y no a 378 (la suma de la segunda columna) porque algunas excepciones lo son de más de una restricción a la vez, es decir, se repiten.

Podemos añadir que todos los órdenes que constituyen excepciones tienen una frecuencia muy baja: el más frecuente (SDV) aparece 43 veces en el corpus, lo que supone una frecuencia relativa de sólo un 0,04%.

Restricción	f.a. err.	f.a. claus.	f.r. err.
0.1	0	99204	0,00
0.2	13	99204	0,01
0.3	0	99204	0,00
0.4	0	99204	0,00
0.5	0	99204	0,00
1	273	99204	0,28
2	87	99204	0,09
3	5	4381	0,11
Total	365	99204	0,37

Cuadro 2: Precisión de las restricciones en la BDS

4.2. Órdenes

Consideramos como órdenes matemáticamente posibles las variaciones simples, es decir, los órdenes resultantes de aplicar las restricciones 0.1, 0.2 y 0.3, de acuerdo con la fórmula 1.

$$V_{m,n} = \frac{m!}{(m-n)!} \quad (1)$$

Si establecemos que las cláusulas que tratamos son compatibles con 7 funciones (m) <S, V, D, I, R, P, O>, que pueden estar formadas por entre una y cuatro de estas funciones (n) y que ninguno de los elementos puede repetirse, el número de órdenes posibles es 1099.

El cuadro 3 muestra la cantidad de órdenes matemáticamente posibles que rechaza cada restricción, o lo que es lo mismo, su capacidad de desambiguación. La cifra que mostramos es el resultado de aplicar cada restricción sobre el total de órdenes teóricamente posibles (que consideramos que son las variaciones simples), independientemente del resto de restricciones.

Restricción	-
0.4	516
0.5	272
1	786
2	636
3	30

Cuadro 3: Resultado de aplicar las restricciones de forma independiente a OM

El cuadro 4, en cambio, muestra cómo se reduce la cantidad de órdenes a medida que aplicamos restricciones, no de forma independiente, sino de forma acumulativa, es decir, sobre el resultado de las restricciones inmediatamente anteriores.

Restricción	+	-
0	1099	
0.4	583	516
0.5	481	102
1	261	220
2	165	96
3	149	16
Total	149	950

Cuadro 4: Resultado de las restricciones aplicadas de forma acumulativa a OM

En la última fila podemos ver que, de 1099 órdenes matemáticamente posibles, las restricciones establecidas han rechazado 950 y por lo tanto quedan sólo 149 órdenes que consideramos como lingüísticamente posibles, y que suponen sólo un 13,56% de los órdenes matemáticamente posibles. Es decir, que poco más de uno de cada diez órdenes matemáticamente posibles son lingüísticamente posibles. Por lo tanto, el orden no es ni mucho menos libre en el sentido literal del término.

En el cuadro 5 mostramos la relación que se establece entre los tres tipos de órdenes (matemáticamente posibles, lingüísticamente posibles y documentados), según el número de elementos que los forman (n), y sin tener en cuenta las excepciones.

n	OM	OL		OD	
	f.a.	f.a.	f.r.	f.a.	f.r.
1	7	1	14,29	1	100
2	42	12	28,57	12	100
3	210	52	24,76	46	88,46
4	840	84	10,00	45	53,57
Total	1099	149	13,56	104	69,80

Cuadro 5: Órdenes M, L y D según el número de elementos que los forman y sin excepciones

En los órdenes formados por uno o dos elementos, lo lingüístico coincide con lo documentado, es decir, que todos los órdenes lingüísticamente posibles están documentados y todos los órdenes documentados son lingüísticamente posibles, y además no se producen excepciones. En cambio, en los órdenes formados por tres y sobre todo por cuatro elementos, se producen ciertos desajustes, ya

que hay órdenes que a pesar de no estar documentados consideramos como lingüísticamente posibles, y hay órdenes que, a pesar de estar documentados, no consideramos como lingüísticamente posibles sino como excepciones, aunque los hayamos excluido del cuadro.

4.3. Restricciones adicionales: los clíticos

Una vez obtenidos los órdenes que consideramos como lingüísticamente posibles, podemos extraer restricciones adicionales sobre la categoría de los argumentos desplazados de su posición prototípica, la aparición de clíticos, etc. En este apartado ilustramos algunas de estas restricciones adicionales, concretamente las relacionadas con los clíticos funcionales: los clíticos de complemento directo e indirecto.

A partir del estudio de la relación entre la presencia de clíticos acusativos y dativos y la presencia y posición del complemento directo e indirecto en la BDS, podemos establecer las siguientes restricciones:

1. Un constituyente antepuesto al predicado que es un pronombre personal precedido por *a* (17) o una cláusula relativa nominalizada, precedida (18) o no (19) por *a*, puede funcionar como complemento directo sólo si hay un clítico acusativo en la misma cláusula.

(17) *A mí*, ni siquiera *me* han detenido todavía.

(18) Si las que estamos dentro nos peleamos, *a quien venga de fuera* nos lo cargamos.

(19) *Lo que yo aún no sepa* te lo enseñará ella. (sonrisa: 322, 3)

2. Si hay un clítico acusativo en la cláusula, un constituyente pospuesto al predicado puede funcionar como complemento directo sólo si es un pronombre personal (20).

(20) Les venderán los ojos y *los* conducirán *a ustedes* a la salida.

3. Un constituyente antepuesto al predicado puede funcionar como complemento indirecto sólo si hay un clítico dativo en la misma cláusula.

(21) *A él* también *le* aconsejaron no dormir en su casa.

Estas restricciones presentan un margen de error de sólo un 0,27% en la BDS.

4.4. Reducción de la ambigüedad

Las restricciones vistas a lo largo del apartado 4 pueden usarse como filtro de la salida de un analizador sintáctico: dado un conjunto de órdenes considerados como posibles por el analizador, como por ejemplo SVCDIC, DSV, PVOC, VDRCO etc.⁹, las restricciones anteriores, ignorando los complementos circunstanciales, seleccionan como orden lingüísticamente posible sólo el primero de ellos: DSV es rechazado por presentar dos funciones argumentales antepuestas, PVOC por contener P y O en la misma cláusula, y VDRCO por contener tres argumentos detrás del predicado.

5. Conclusiones

Hemos propuesto un método basado en corpus para la descripción del orden de constituyentes del español. Hemos aplicado este método a las cláusulas más frecuentes en español (cláusulas activas personales independientes o subordinadas introducidas por conjunción) y a un subconjunto de las funciones que puede contener una cláusula (las funciones argumentales más el predicado, que forman el esquema verbal) y hemos extraído algunas de las restricciones que las caracterizan. La principal ventaja de este enfoque es que se basa en datos reales, extraídos de corpus, y que aspira a describir el orden desde un punto de vista formal, para contribuir a la desambiguación sintáctica, que es un problema esencial en el análisis sintáctico automático. Podemos establecer los siguientes objetivos futuros:

- La descripción del orden del resto de tipos de cláusulas: impersonales, pasivas, no finitas, interrogativas y relativas.
- La extensión de la tasa de error para captar más restricciones
- El establecimiento de una jerarquía entre las restricciones: algunas de ellas afectan a todos los tipos de cláusulas y otras son específicas de un tipo.
- La aplicación de las restricciones a la mejora de un sistema de análisis sintáctico

automático y el estudio de cuanta ambigüedad reduce la información sobre el orden.

- La extracción de restricciones que puedan afectar a los complementos circunstanciales.

Bibliografía

- Butt, J. y C. Benjamin. 1988. Word order. En *A new reference grammar of modern spanish*. Edward Arnold, London, capítulo 31, páginas 394–400.
- Civit, M. y M.A. Martí. 2002. Design Principles for a Spanish Treebank. En *Proceedings of the First Workshop on Treebanks and Linguistic Theories*, páginas 61–77.
- Delbecque, N. 1991. *El orden de los sintagmas. La posición del regente*. Universidad de Salamanca, Salamanca.
- Fernández Ramírez, S. 1951. El orden de las palabras: la posición del sujeto. En I. Bosque, editor, *Gramática española (4): El verbo y la oración*. Arco libros, Madrid, capítulo 9, páginas 430–462.
- Fish, G.T. 1960. Posverbal word order in spanish prose. *Hispania*, 43:426–429.
- Gala, N. 1999. Using incremental finite-state architecture to create a spanish shallow parser. *Procesamiento del Lenguaje Natural*, 25:75–82.
- Gili Gaya, S. 1961. Orden de colocación de los elementos oracionales. En *Curso superior de sintaxis española*. Bibliograf, Madrid, capítulo VI, páginas 81–94.
- Gutiérrez, M.L. 1978. *Estructuras sintácticas del español actual*. SGEL, Madrid.
- Holan, T., V. Kuboň, K. Oliva, y M. Plátek. 2000. On complexity of word order. *TAL. Traitement automatique des langues*, 41(1):273–300.
- López García, A. 1996. La informatividad. En López García, A., editor, *Gramática del español II. La oración simple*. Arco libros, Madrid, capítulo 21, páginas 449–510.
- Pineda, L. y I. Meza. 2003. Una gramática básica del español en HPSG. Informe técnico, Departamento de Ciencias de la Computación, IIMAS, UNAM, México.

⁹Donde C es “complemento circunstancial”.

- Rojo, G. 2001. La explotación de la Base de datos sintácticos del español actual (BDS). En J. de Kock, editor, *Lingüística con corpus. Catorce aplicaciones sobre el español*. Universidad de Salamanca, Salamanca, páginas 255–286.
- Santana, O., J.R. Pérez, J.L. Losada, y F.J. Carreras. 2004. Bases para la desambiguación estructural de árboles de representación sintáctica. *Procesamiento del Lenguaje Natural*, 32:43–65.
- Torrego, E. 1984. On inversion in spanish and some of its effects. *Linguistic Inquiry*, 15(1):103–129.
- Valverde, M.P., M.P. Santalla, y M. Civit. 2005. Constituent order in spanish: a syntactically annotated corpus-based study. En *Proceedings of the Fourth Workshop on Treebanks and Linguistic Theories*, páginas 197–208.

A. Anexo 1: Siglas

A.1. Categorías

sn:det.nc="sintagma nominal formado por un determinante y un nombre común (que llevar complementos)"

- det = "determinante"
- nc: = "nombre común"
- np="nombre propio"
- x ="sintagma preposicional, adjetivo o pronombre"
- pp="pronombre personal"
- pin="pronombre indefinido o numeral"
- pd="pronombre demostrativo"
- px="pronombre posesivo"
- S.F.C.que="subordinada sustantiva introducida por *que*"
- S.F.C.Q="subordinada interrogativa indirecta parcial"
- S.F.C.si="subordinada sustantiva introducida por *si*"
- S.F.A="subordinada adverbial"
- S.NF.C="subordinada de infinitivo"
- S.NF.A="subordinada de gerundio"
- S.NF.P="subordinada de participio"
- S.F.R="subordinada relativa"

- sadj="sintagma adjetival"
- sadv="sintagma adverbial"
- p="preposición"

A.2. Funciones

- S="sujeto"
- V="predicado"
- D="complemento directo"
- I="complemento indirecto"
- R="complemento de régimen"
- P="complemento predicativo del sujeto"
- O="complemento predicativo del complemento directo"
- A="complemento agente"