

Caracterización sintáctica de los verbos castellanos para una gramática en ordenador

**Consuelo Rodríguez Margo, Luis Sopeña, Celia Villar
Centro de Investigación UAM-IBM
Madrid**

Resumen

El presente artículo muestra un intento de caracterización sintáctica de los verbos castellanos, uno de cuyos objetivos es la realización de una gramática en ordenador. Para ello, se ha trabajado con códigos de dos tipos: alfabéticos y numéricos.

RESUMEN

El presente artículo muestra un intento de caracterización sintáctica de los verbos castellanos, uno de cuyos objetivos es la realización de una gramática en ordenador. Para ello, se ha trabajado con códigos de dos tipos: alfabéticos y numéricos.

INTRODUCCIÓN

La realización de una gramática supone un estudio de la lengua en todos sus niveles: morfológico, sintáctico y semántico. En nuestro caso, se trata de una gramática en ordenador y, necesariamente, hay que abordar el tema, en principio, desde los mismos presupuestos, para dar cuenta de los mismos hechos, aunque la metodología sea diferente.

Para trabajar en sintaxis será necesario estudiar el comportamiento de las palabras de la lengua en cuestión, dar cuenta del uso que se hace de las mismas. Este artículo presenta una tentativa de caracterización sintáctica de los verbos castellanos, que supone, para nosotros, un paso más de estudio en un nivel superior de la lengua. El trabajo que en él se expone se integra en un entorno más amplio, en el que se han construido una serie de herramientas para la composición y verificación de documentos escritos en castellano, como un diccionario de verificación ortográfica y otro de sinónimos, un analizador morfológico, etc. [1]

Tal intento de caracterización ha consistido en la asignación de una serie de códigos gramaticales, diseñados para unos 4.000 verbos [2], que refleja su comportamiento sintáctico: tipo de complementos que puedan llevar (complementos preposicionales, un infinitivo, un gerundio...), régimen preposicional de los mismos, etc. Hay que señalar la importancia que un trabajo de estas características tiene para el castellano, puesto que, hasta el momento, existen pocos estudios relativos al comportamiento sintáctico de los verbos [3]. Para el inglés existe, por ejemplo, una fuente de consulta muy valiosa: *The Longman Dictionary of Contemporary English* [4], en el que cada entrada lleva asociada información gramatical que intenta describir el comportamiento sintáctico. Este diccionario utiliza criterios e información razonables, que nos han sido de ayuda en un primer momento, pero, puesto que contamos con estructuras particulares que nos diferencian del inglés, hemos diseñado los códigos que se adaptan a nuestra lengua.

Para el castellano no hay un diccionario similar, y los que existen sólo presentan el material de forma convencional: entrada, parte de la oración, definición, acepciones... Por lo tanto, es necesario contar con un diccionario más completo, que haga una descripción más precisa de la lengua, si queremos elaborar una gramática que sirva como base para posteriores estudios (análisis sintáctico, traducción automática, etc.).

LOS CÓDIGOS SINTÁCTICOS

Con los objetivos mencionados, y partiendo de ese corpus aproximado de 4.000 verbos, hemos creado unos códigos alfabéticos y numéricos que describimos a continuación. Los primeros definen el tipo de verbo y los segundos la clase de complementos que admiten o exigen.

1. Códigos alfabéticos

- T, verbos transitivos, exigen un objeto directo
(ej. *acusar, terminar*)
- I, verbos intransitivos, no llevan objeto
(ej. *temblar, tardar*)
- verbos ditransitivos, que comprenden tres casos:
 - DI, los que admiten objeto directo e indirecto
(ej. *regalar algo a alguien*)
 - DD, los que admiten dos objetos directos
(ej. *preguntar algo a alguien*)
 - DC, los que admiten un objeto directo y un complemento del objeto
(ej. *considerar algo (predicativo) a alguien (obj. directo)*).
- A, verbos auxiliares. En este grupo no sólo están los clásicos como 'ser' o 'estar', sino también aquellos verbos que se han gramaticalizado en alguno de sus usos, es decir, al unirse a las formas no personales del verbo (infinitivo, gerundio o participio), que es la que lleva la carga semántica. Adquieren una función parecida a la del verbo 'haber', por ejemplo, en la formación de los tiempos compuestos, (ej. *dejó escrita su última voluntad*).
- TP, verbos terciopersonales. Son verbos sin sujeto (ej. *hace frío*).
- verbos pronominales, con dos subgrupos:
 - PS, los que pueden llevar, variando de significado, un pronombre referido al sujeto en toda la conjugación
(ej. *acordar (algo)/acordarse (de algo)*)
 - PP, los que únicamente son pronominales,
(ej. *arrepentirse, atreverse*)
- verbos que sólo se usan con el pronombre 'SE', para lo cual nos hemos inspirado, entre otros, en los criterios expuestos en [5].
 - RT, dicha forma pronominal aparece
 - * con un verbo transitivo, y existe concordancia entre el verbo y el elemento nominal de la oración (es lo que se llama **pasiva pronominal**)
(ej. *el café SE CULTIVA en los países tropicales*)
(ej. *los plátanos SE CULTIVAN en...*)

* con un verbo transitivo, y no existe concordancia entre el verbo y el elemento nominal de la oración (oración impersonal)
(ej. *SE AVISA a los interesados*)

- RS, con expresión de un sujeto activo
(ej. *SE CONFIRMÓ la noticia por su padre*)

- RI, con verbo intransitivo
(ej. *SE VIVE bien de vacaciones*)

- RA, con verbo atributivo
(ej. *SE PUEDE SER pobre y feliz*)

A pesar de esta clasificación, no hay que perder de vista que, más que de verbos transitivos, auxiliares..., habría que hablar de construcciones de tal o cual naturaleza, puesto que está claro que un mismo verbo puede presentarse de muchas formas, en distintas estructuras y que, en último término, la estructura del predicado depende de la cantidad de información que queramos dar. Por eso, hemos intentado recoger los usos y acepciones de cada verbo, y, en consecuencia, la mayoría de ellos llevarán asignados varios códigos.

2. Códigos numéricos

- 0, para verbos que admitan un adverbio, una frase preposicional como complemento o que no lleven complemento.
(ej. *vino A LAS TRES*)
- 1, para verbos que admitan un nombre, pronombre o frase nominal entre sus complementos.
(ej. *cómprale un vestido, LA acusaron de asesinato, adoptaron MEDIDAS DRÁSTICAS*)
- 2, por el momento no se usa
- 3, verbos que admiten un infinitivo como complemento. Se especificará si el infinitivo puede aparecer o no con alguna preposición por delante (y se especifica ésta).
(ej. *se encargó DE recoger el paquete*), con un código PP3 (de)
(ej. *supo callar a tiempo*)
- 4, verbos que admiten un gerundio entre sus complementos.
(ej. *salió CORRIENDO, anda BUSCANDO empleo*)
- 5, verbos que introducen una oración completiva introducida con "QUE".
(ej. *quiero QUE TE VAYAS, me apetece QUE ME ACOMPAÑES*)
- 6, verbos que introducen una oración interrogativa.
(ej. *dime QUIÉN HA VENIDO, calcula CUÁNTO CUESTA*)
- 7, verbos que llevan un adjetivo como complemento.
(ej.: *hpl.salió DEPRIMIDO de la reunión, se confesó CULPABLE*)
- 8, para verbos seguidos del participio pasado de otro verbo.
(ej. *dejó ESCRITO lo que pensaba hacer, ya lleva LEÍDOS dos libros*)

Así pues, cada verbo va a estar caracterizado por una colección de pares código alfabético-código numérico, que intentará cubrir toda la gama de posibilidades de uso de dicho verbo.

EJEMPLOS

A continuación presentamos unos ejemplos de códigos verbales. Con cada estructura diferente hay una combinación de los dos tipos de códigos, el alfabético y el numérico.

acordar

T1 Acordaron la suspensión de la reunión
T3 Acordaron suspender la reunión
T5 Acordaron que la reunión se suspendiera
PS0 Me acuerdo perfectamente de su cara
PS3 Me he acordado de recogerlo
PS5 Me acuerdo de que venía todos los días

deber

T1 Debe mil pesetas
DI1 Me debe mil pesetas
A3 Debo llegar a las tres (= 'tengo que llegar a las tres')
A3 Dēben de ser las tres (= 'puede que sean las tres')
PS0 Me debo a mi familia

decir

T1 Dijo la verdad
T5 Dijo que se iba
T6 Dijo quién era el vencedor
D1 Le dijo la verdad
D5 Nos dijo que se iba
D6 Nos dijo quién era el vencedor
IO No me digas más
RT1 Se dicen tantas cosas
RT5 Se dice que es viuda

dejar

T1 Deja a su familia en la miseria
T5 Deja que se vaya
D7 Dejó estupefactos a los asistentes
PS1 Me he dejado las gafas en la mesa
PS3 Me dejo llevar por las circunstancias
A3 Deja correr el agua / Deja de gritar
A8 Dejó escrito lo que pensaba hacer

APLICACIÓN

Este trabajo, de extraordinario interés para el castellano, tendrá que completarse con un estudio similar para el resto de las palabras.

Hasta ahora, hemos hablado de un esbozo de los códigos. La primera aplicación a la que se destina este estudio de los verbos es la elaboración de una gramática del castellano, ya iniciada. Se está escribiendo en el lenguaje PLNLP [6] y, para hacer los códigos más adecuados a su uso, es más conveniente que éstos, que caracterizan a los verbos de una forma demasiado precisa, se traduzcan en unos rasgos más flexibles que engloben a los dos tipos de códigos de los que hemos hablado. Así, el código T se convierte en un rasgo TRAN; DI en DITRAN; 5 en THATCOMP, etc.

Los rasgos se utilizarán en reglas del siguiente tipo:

VP (...TRAN...) NP ---> VP (... ,OBJDIR=NP,...)

Por otro lado, al analizar oraciones sintácticamente, la gramática tendrá que completarse, añadiendo, eliminando o cambiando los rasgos que definan a las palabras. Además, será necesario completarla con rasgos de tipo semántico que ayuden a la detección de frases semánticamente anómalas, aunque sintácticamente correctas. Éste será uno de los objetivos de la siguiente fase de nuestro trabajo.

REFERENCIAS

- [1] Casajuana, R., Rodríguez, C., Sopena, L., Villar, C., Towards an Integrated Environment for Spanish Document Verification and Composition, Proceedings 3rd. Conference Association for Computational Linguistics (European Chapter), Copenhagen, April, 1987.
- [2] Casajuana, R., Rodríguez, C., Clasificación de los verbos castellanos para un diccionario en ordenador. Actas del I Congreso de Lenguajes Naturales y Lenguajes Formales, Univ. de Barcelona, Fac. de Filología, Dep. de Lingüística General, Octubre 1985.
- [3] Subirats-Rüggeberg, C., Sentential Complementation in Spanish. "A lexico-grammatical study of three classes of verbs". *Lingvisticae Investigationes: Supplementa*. Volume 14. John Benjamins Publishing Company, Amsterdam/Philadelphia, 1987.
- [4] Longman Group UK Limited, England. Reprinted 1986.
- [5] Moliner, M., Diccionario de uso del español. Ed. Gredos, Madrid, 1982.
- [6] Heidorn, G., Augmented Phrase Structure Grammars, in Nash-Webber and Schank (eds.), *Theoretical Issues in Natural Language Processing*, ACL, 1975.