

LAS CONSTRUCCIONES COMPARATIVAS EN CORPUS DEL ESPAÑOL SINTÁCTICAMENTE ANOTADOS

M.^a Paula Santalla del Río/Eva M.^a Domínguez Noya
Universidad de Santiago de Compostela/Centro Ramón Piñeiro
para a Investigación en Humanidades

<http://dx.doi.org/10.18778/8220-201-4.26>

Resumen

En esta exposición, tras una breve explicación de los presupuestos en que se basa la elaboración del corpus CSA y de lo que entendemos que subyace a los niveles de descripción gramatical que pueden reconocerse, se aborda el estudio de las estructuras comparativas, clasificadas en grupos a partir fundamentalmente de la *Nueva Gramática de la Lengua Española (NGLE)*, en cuanto a como puede variar y varía su análisis en tres corpus sintácticamente analizados (*UAM Spanish Treebank*, *IULA Spanish LSP Treebank* y *AnCora-es*) y con respecto especialmente al propio corpus CSA. De ello se concluye que, para los propósitos de la anotación de corpus y la reutilización de recursos, una reflexión conjunta que camine hacia una estandarización en el paso de un nivel de descripción no formalizado a uno que sí lo esté se hace hoy necesaria.

Palabras clave: Corpus, anotación sintáctica, estructuras comparativas, formalización.

1. Introducción

La construcción de recursos lingüísticos, entre los que se encuentran los corpus o bases de datos textuales, es fundamental en toda lengua para el avance en el procesamiento automático del lenguaje

natural, pero además es también especialmente beneficiosa para profundizar en el conocimiento mismo de la lengua, porque obliga a una descripción formalizada del material analizado. La creación de CSA, un corpus sintácticamente anotado del español¹, se concentró en este hecho, dado que su objetivo primordial fue la exploración en detalle del instrumento que codifica el análisis del recurso en sí. Este planteamiento nació de la observación de las repercusiones que en la elaboración de los corpus tiene la distinción entre distintos niveles de descripción gramatical: descripción gramatical, descripción gramatical comprensiva y exhaustiva, y descripción gramatical formalizada; este último es el necesario para analizar corpus lingüísticos, pero es un nivel descriptivo que cada corpus anotado suele en la práctica alcanzar de manera independiente de los demás.

Como resultado de lo anterior, se observa que la heterogeneidad de los análisis recogidos en los corpus para las mismas estructuras es con frecuencia muy amplia, lo cual dificulta enormemente la reutilización posterior de unos recursos cuyo desarrollo es muy costoso. Tal heterogeneidad es especialmente apreciable en estructuras sobre cuya descripción no hay ya de partida un acuerdo tan consensuado como para otras.

A la luz de la diversidad de lo que se ha dicho de ellas, las construcciones comparativas se encuentran, precisamente, entre las que presentan mayor complejidad sintáctica de toda la gramática española. En este trabajo a partir de una clasificación de estas estructuras derivada de una serie de estudios sobre ellas (Gutiérrez, 1994; Sáez, 1999; Brucart, 2003; NGLÉ, 2009; Sáez, Sánchez, 2014), revisaremos cómo tales estructuras han sido tratadas en otros corpus analizados sintácticamente (*el UAM Spanish Treebank* [<http://www.llf.uam.es/ESP/Treebank.html>, Moreno *et al.*, 1999], *IULA Spanish LSP Treebank* [<https://repositori.upf.edu/handle/10230/20048>, Marimon *et al.*, 2012] y *AnCora-es* [<http://clic.ub.edu/corpus/es>, Martí *et al.*, 2008]), y compararemos esos tratamientos con la presentación de cómo el sistema de anotación de CSA analiza las comparativas, todo ello con el fin de mostrar

¹ <http://gramatica.usc.es/proyectos/drasae/>.

como, para los propósitos de la Lingüística de corpus y el Procesamiento del Lenguaje Natural, parece hoy, efectivamente, hacerse más necesario que nunca avanzar en la propuesta de estándares de descripción formal y análisis de corpus. No podemos, pues, conformarnos en el momento actual con la difusión que de modo natural han alcanzado estándares que lo son fundamentalmente de formato, como el *CoNLL*, sino que se hace preciso proponer y promocionar estándares de análisis descriptivo, para en un futuro próximo disponer de recursos plenamente reutilizables e integrables, así como producir masivamente datos que puedan servir para alimentar modelos de análisis lingüístico automático.

Tras una breve explicación de los presupuestos en que se basa la elaboración del corpus CSA (apartado 2) y de lo que subyace a los niveles de descripción referidos (apartado 3), en el apartado 4, a partir de una clasificación de las estructuras comparativas basada fundamentalmente en la *Nueva Gramática de la Lengua Española (NGLE)*, pero también en otros autores (Gutiérrez, 1994; Sáez, 1999; Sáez y Sánchez, 2014), se aborda su estudio en cuanto a cómo varía su análisis en tres corpus sintácticamente analizados (*UAM Spanish Treebank*, *IULA Spanish LSP Treebank* y *AnCora-es*) y con respecto especialmente al propio corpus CSA. Tras todo ello, se exponen las conclusiones que cabe extraer de esta revisión.

2. El corpus CSA

El corpus CSA² contiene, por el momento, el análisis completo, llevado a cabo manualmente, de una única obra: *Crónica de una muerte anunciada*, de Gabriel García Márquez. La anotación realizada es exclusivamente de tipo sintáctico; esto es, segmentación y asignación exhaustivas de función sintáctica y tipo de unidad a todos los segmentos identificados para cada secuencia aislada que forma parte del corpus.

² Se puede consultar y descargar en <https://galvan.usc.es/drasae/>.

2.1.

El sistema de anotación³

Para la elaboración del sistema de anotación de CSA fueron consultados los criterios y sistemas de anotación de los ya no pocos, y para no pocas lenguas, corpus sintácticamente analizados existentes. No hay lugar aquí para una revisión de todos ellos: nos referimos especialmente a los corpus *AnCora*, para español y catalán (Taulé *et al.*, 2008), y, a partir de ellos, a los que los inspiraron, citados en Martí *et al.* (2008).

De manera breve, el sistema de anotación aplicado a CSA puede ser descrito a partir de las siguientes afirmaciones: a) codifica un análisis constitutivo y no dependencial, b) se sustenta en un modelo teórico constitutivo y funcionalista, c) es exhaustivo en cuanto al reconocimiento de constituyentes y la asignación a los mismos de funciones y unidades sintácticas, d) respeta estrictamente el carácter presencial y el orden secuencial de los constituyentes reconocidos, y e) siendo consistente en la anotación del mismo hecho o de hechos parcialmente semejantes a lo largo del corpus, es lo más simple posible y se propone resultar familiar a sus posibles usuarios recurriendo a las tradiciones descriptivas más ampliamente reconocidas (*NGLE*).

Con tales premisas, CSA se analiza con un *tagset* de 42 etiquetas funcionales y 32 estructurales, apoyadas por tres recursos adicionales de representación: sangrado y/o dígitos de nivel (para la representación de la jerarquización), coindización de elementos al margen de la jerarquización (para elementos funcionales más allá del segmento en el que se analizan) y numeración romana de constituyentes interrumpidos por otros constituyentes al mismo nivel⁴.

Así, en CSA cada secuencia analizada es un árbol tradicional, solo que dispuesto verticalmente. En cada línea horizontal se identifica un segmento resultado de la progresiva segmentación de la secuencia analizada, y se le asigna a ese segmento una unidad y una función. Sangrado y un dígito que identifica el nivel de

³ Para más detalles remitimos a lo ya presentado a este respecto en Santalla (2012).

⁴ Se puede consultar en <https://gramatica.usc.es/wiki/drasae>.

profundidad alcanzado hasta ese punto en el análisis representan la jerarquización. La visualización del análisis de una secuencia en la aplicación *PaME* es la que vemos en la figura 1: muy amigable para el usuario y en la que solo los atributos del PREDICADO quedan ocultos, a la espera de que se pose sobre dicha función sintáctica el puntero del ratón para que emerja un cuadro con sus características:

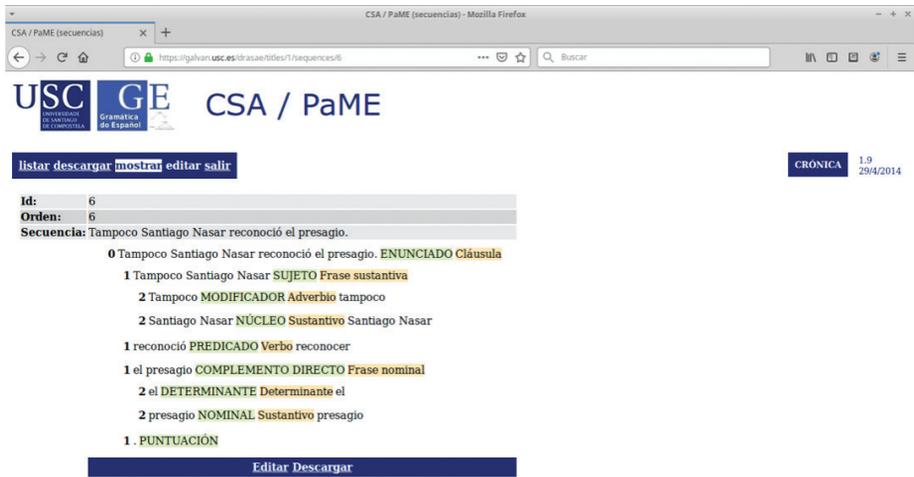


Figura 1. Análisis en CSA de *Tampoco Santiago Nasar reconoció el presagio*

3. La descripción lingüística y la anotación lingüística

Aunque la segunda de estas operaciones, asociada al desarrollo de corpus lingüísticos en formato electrónico y de reciente aparición, se lleva a cabo sobre la base de fundamentos que ha sentado la primera a lo largo de su ya prolongada historia, una y otra son operaciones muy distintas, cuyas respectivas prácticas de trabajo son en consecuencia también diferentes. Esta diferencia, unida a la circunstancia de que la descripción a la que pueden acceder quienes anotan no se proporciona siempre en el nivel de desarrollo por

esta operación requerido, está en la base de la heterogeneidad de los variados análisis codificados por los anotadores para las mismas estructuras en distintos corpus lingüísticos. Estos hechos, que comprobaremos para las estructuras comparativas en el apartado sucesivo, se derivan de los condicionamientos, relacionados con el contexto en el que se desenvuelven esos procesos, que determinan las actividades de descripción y anotación lingüística.

Así, mientras la anotación es una actividad que se enfrenta a texto real y ha de analizarlo exhaustiva y secuencialmente, la descripción es una operación que, sobre la base de una clasificación previa de los hechos lingüísticos, se focaliza sucesivamente en los que describe, haciéndolo de tal manera que incluso los ejemplos con los que se ilustran esos hechos, que no siempre son de texto real, se describen solo con respecto a los que se pretende que ilustren (sin describir lo que tienen a su alrededor)⁵.

Por otro lado, además de lo anterior, tenemos que tener presente que, contemplada globalmente, la descripción lingüística en la que se basan quienes anotan puede encontrarse en distintos niveles de desarrollo. Puede ser un conjunto de descripciones parciales, en documentos distintos y cada una sobre un fenómeno concreto, que se analiza más o menos en profundidad (con suerte utilizando presupuestos semejantes), o puede ser una descripción comprensiva, general y en un único documento, que analiza cada fenómeno con menor detalle, pero más homogéneamente en relación con otros fenómenos considerados.

El estadio de desarrollo en el que se cuenta con gramáticas descriptivas comprensivas y detalladas (como la *NGLE*) facilita enormemente el trabajo de diseño y aplicación de un sistema de anotación, pero se necesita ir aún más allá para, incluso sobre la base de estas descripciones, anotar corpus lingüísticos: la descripción tiene que estar formalizada, esto es, tiene que haber sido reducida

5 Una muestra de la insuficiencia con la que la descripción sirve a la anotación es la falta de atención que la primera presta a la integración, ineludible en la anotación, de lo discursivo en lo sintáctico (¿dónde se ubica, por ejemplo, en el análisis sintáctico codificado en un corpus un conector discursivo?).

a una notación limitada de etiquetas categoriales y/o funcionales representable gráficamente por medio de unos recursos también limitados.

Aparte del hecho de que pueden remitirse a marcos teórico-descriptivos distintos (algo que es difícil de salvar), incluso cuando se remiten al mismo, del hecho de que cada grupo que desarrolla un corpus formaliza independientemente de los demás la descripción más o menos tradicional de la que parte, surgen las diferencias entre corpus anotados, y de manera más acusada, por la más amplia naturaleza de los fenómenos tratados, las diferencias entre corpus sintácticamente anotados. Ilustramos esas diferencias en el apartado siguiente, a partir del análisis que de las estructuras comparativas algunos corpus sintácticamente analizados y especialmente el corpus CSA han codificado, basándose para ello en tradiciones descriptivas que se reconocen como comparables.

4. Las estructuras comparativas

Las construcciones comparativas están, en opinión de muchos especialistas, entre las que presentan mayor complejidad sintáctica de toda la gramática española. No todos los gramáticos reconocen, al menos explícitamente, como comparativas el mismo conjunto de estructuras, pero todos consideran tales al menos, de entre las que se han llamado tradicionalmente de grado, a las constituidas en torno a *más/menos... que/de* y *tanto/a/os/as/tan... como*. En ellas nos centramos a continuación.

4.1. Comparativas de grado

Si bien, *grosso modo*, hay acuerdo a la hora de clasificarlas semánticamente en función de los cuantificadores que las componen, distinguiendo entre comparativas de igualdad, superioridad

e inferioridad (estos dos últimos tipos englobados en las comparativas de desigualdad), o en función de la magnitud comparada, según la cual, dependiendo de la clase de palabra a la que modifique el cuantificador comparativo, se establece la existencia de comparativas de grado cualitativas (con adjetivos y adverbios) y comparativas de grado cuantitativas (con sustantivos), incluso con respecto a este primer grupo de comparativas, sigue sin haber todavía una propuesta de análisis sintáctico integral que dé cuenta de ellas.

Es cierto que, con algunas diferencias relativas a elementos más bien secundarios, reconocidos o no por unos u otros, y, sobre todo, con diferencias terminológicas, más o menos todos los gramáticos reconocen en estas comparativas (en términos de la *NGLE*) un primer término de la comparación, un grupo cuantificativo y un segundo término de la comparación o complemento comparativo. Pero también es cierto que ninguno de esos gramáticos alcanza un consenso general con el resto, más allá de ese punto de la descripción, sobre detalles ulteriores de la misma. Y menos aún se produce el consenso en cuanto a la formalización, que ni siquiera se plantean en la mayor parte de los casos, de las estructuras en cuestión.

¿Cómo hemos formalizado nosotros ese consenso descriptivo, y lo que va más allá de él, en la anotación de *CSA*? Seleccionamos, para los propósitos de esta exposición, tres ejes de variación observados en las descripciones revisadas y en los corpus examinados, y vemos primero que en *CSA* los tratamos, como se evidencia en el análisis de la figura 2, del modo siguiente: i) reconocemos como estructura comparativa (con la etiqueta correspondiente) al grupo cuantificador y al segundo término, dejando fuera el primero; ii) reconocemos una posible construcción verbal (una cláusula) en el segundo término; iii) entendemos que el segundo término modifica al grupo cuantificativo y no solo al cuantificador.

- 0 Ella no estuvo entonces tan segura como la primera vez. ENUNCIADO Cláusula
- 1 Ella SUJETO Pronombre ella
- 1 no estuvo PREDICADO Verbo estar
- 1 entonces COMPLEMENTO CIRCUNSTANCIAL Adverbio entonces
- 1 tan segura como la primera vez COMPLEMENTO PREDICATIVO Estructura comparativa
- 2 tan segura COMPARADOR Frase adjetiva
- 3 tan MODIFICADOR Adverbio tan
- 3 segura NÚCLEO Adjetivo seguro
- 2 como la primera vez 2.º TÉRMINO Fragmento
- 3 como COMPLEMENTO CIRCUNSTANCIAL Adverbio como
- 3 la primera vez COMPLEMENTO CIRCUNSTANCIAL Frase nominal
- 4 la DETERMINANTE Determinante el
- 4 primera vez NOMINAL Frase nominal
- 5 primera DETERMINANTE Determinante primer
- 5 vez NOMINAL Sustantivo vez
1. PUNTUACIÓN

Figura 2. Análisis en CSA de *Ella no estuvo entonces tan segura como la primera vez*

Así, respecto a i) la **Estructura comparativa** (utilizamos ahora las denominaciones de las etiquetas de CSA) abarca al segmento *tan segura como la primera vez* y deja fuera al primer término de la comparación (*entonces*). Con respecto a ii), el **2.º TÉRMINO** está constituido por un **Fragmento**, una denominación que corresponde, en este caso, a una cláusula (construcción verbal) en la que falta el verbo (pero en la que seguimos teniendo dos funciones clausales, dos instancias de **COMPLEMENTO CIRCUNSTANCIAL**). Con respecto a iii), el **2.º TÉRMINO** se encuentra al nivel jerárquico del segmento *tan segura*, con el que constituye la **Estructura comparativa**, y no al de *tan*, en conjunción con el cual se encontraría al nivel de *segura* y constituiría la **Estructura comparativa** en un análisis que ligara el segundo término al cuantificador y no al grupo cuantificativo.

¿Qué sucede, en cambio, en otros corpus sintácticamente analizados? En el corpus *UAM Spanish Treebank* se identifican estructuras comparativas de esta clase (no siempre del mismo modo) con los elementos comparativos grupo cuantificativo y segundo término, y con un segundo término que se relaciona con el grupo cuantificativo, pero no se reconoce consistentemente el carácter de construcción verbal de ese segundo término. Así en *Y augura el advenimiento de una era de ordenadores [que]_{1.º térm.}*

posiblemente un día sean [miles de millones de veces más rápidos]_{GC} [que los superordenadores más potentes de hoy]_{2.º término}, no se reconoce una función sujeto asociada a los superordenadores más potentes de hoy.

En el corpus *AnCora* no se identifican estructuras comparativas como tales, pero se reconoce el carácter de construcción verbal del término con *que*, que se relaciona con el grupo cuantificativo. Así en *El volumen exportado ascendió a 74 millones de litros, [un 22,17 por ciento más]_{GC} [que el año pasado, cuando se vendieron al exterior 60.57 millones, según precisó un informe de la Asociación de Productores y Exportadores de Vinos Finos (Chilevid)]_{2.º término}*, no se reconoce una estructura comparativa asociada a *un 22,17 por ciento más que el año pasado [...]* (*Chilevid*).

En el corpus *IULA Spanish LSP Treebank*, por fin, no se identifican estructuras comparativas como tales, y no se reconoce el carácter de construcción verbal del término con *que*, que se pone en relación con la que sería la cabeza del grupo cuantificativo. Así en *[La papila]_{1.º término} tiene una coloración [más pálida]_{GC} [que el fondo]_{2.º término}*, no se reconoce una estructura comparativa asociada a *más pálida que el fondo*, que se analiza exactamente igual a como se analizaría *muy llena de vida*.

4.2.

Otras comparativas de grado

Nos referimos aquí a las comparativas con *que/como* iguales a las anteriores, pero de alcance frasal (esto es, que no pueden acceder a un primer término de comparación más allá de la frase que constituye su grupo cuantificativo: *Compró unos libros más caros que los que compraste tú/Compró unos libros más caros que interesantes/*Compró unos libros más caros que tú*), y a las comparativas con *de + relativa* (preferentemente, pero también otras unidades), de alcance oracional (*Compra más libros de los que puede leer*) o frasal (*Compró unos libros más interesantes de lo predecible/de lo que de él cabía esperar*).

De la misma manera que las anteriores, las comparativas de grado de estas clases son reconocidas como tales en *CSA*, sin que,

obviamente, se reconozca una estructura clausal en el segundo término de la comparación. Veámoslo con el ejemplo⁶:

2 más serio de lo que hacían creer sus travesuras **MIEMBRO Estructura comparativa**
 3 más serio **COMPARADOR Frase adjetiva**
 4 más **MODIFICADOR Adverbio más**
 4 serio **NÚCLEO Adjetivo serio**
 3 de lo que hacían creer sus travesuras **2.º TÉRMINO Frase preposicional**
 4 de **DIRECTOR Preposición de**
 4 lo que hacían creer sus travesuras **TÉRMINO Cláusula**

Figura 3. Análisis en CSA de *más serio de lo que hacían creer sus aventuras*

Se reconoce un **2.º TÉRMINO**, que es una **Frase preposicional**, *de lo que hacían creer sus travesuras*, que se relaciona en el interior de una **Estructura comparativa** con el **COMPARADOR más serio**, que es el grupo cuantificativo.

¿Qué sucede al respecto en los otros corpus que hemos revisado? Lo mismo que sucedía con las comparativas de grado prototípicas de alcance oracional sucede con las de alcance frasal (evidentemente sin reconocimiento, donde antes lo hubiera, de estructura verbal en el segundo término). Por lo que se refiere a las comparativas de grado con *de*, las otras en las que nos fijábamos en este apartado, también en el corpus *UAM Spanish Treebank* son reconocidas como tales si el segundo término es una relativa, no si es otra unidad (no al menos consistentemente):

Sin embargo, cuando se trata de una auténtica metedura de pata, hay dos posibilidades: errores a nuestro favor (hemos pagado menos o Hacienda nos ha devuelto [[más]GC]1.º término. [de lo que nos corresponde]2.º término.) o errores en contra (hemos

⁶ Por razones de espacio, reducimos a partir de aquí la presentación de los análisis a la estructura comparativa allí donde se encuentre (en este caso en *Me pareció más serio de lo que hacían creer sus travesuras, y de una tensión recóndita apenas disimulada por sus gracias excesivas*) y a los niveles de análisis relevantes en su interior.

pagado más o nos han devuelto [[menos]GC]1.º término. [de lo debido]2.º término.); *El ucranio Honchar gana una contrarreloj en la que Clavero perdió* [[más tiempo]GC]1.º término. [del previsto]2.º término. No se reconoce en estos ejemplos la estructura comparativa en *más tiempo del previsto y sí en más de lo debido* y en *más de lo que corresponde*.

Las comparativas de grado con *de* no son reconocidas como tales en el corpus *AnCora*. Así en

Sorprendente esa actitud activa en un hombre que está alcanzando en esta Vuelta mucho más de lo que nadie, excepto quizá él, podía presumir o en *En Teherán no habrá segunda vuelta electoral, pues 32 candidatos, dos más de los necesarios, han conseguido más del 25 por ciento de los votos requeridos para ser designados en la primera ronda [...]*, no se reconocen comparativas ni en *mucho más de lo que nadie, excepto quizá él, podía presumir*, ni en *dos más de los necesarios* ni en *más del 25 por ciento de los votos requeridos [...]*.

Las comparativas de grado con *de*, por fin, no son reconocidas como tales en el corpus *IULA Spanish LSP Treebank*. Así no se reconocen estructuras comparativas ni en *más de lo que creemos* en *A veces sabemos más de lo que creemos* ni en *más de lo necesario* en *La reunión se alargaba más de lo necesario* y *los ánimos comenzaban a encrespase*.

4.3. Comparativas proporcionales

Las comparativas proporcionales (correlativas, introducidas por *cuanto/os/as/a más/menos...* en un primer miembro y *más/menos...* en un segundo) son reconocidas como tales en *CSA*. En el análisis de *cuanto más aumentaban los planes de la fiesta, más ideas de delirio se le ocurrían para hacerla más grande* (cuyo tamaño no nos permite reproducirlo aquí), vemos que, en atención al hecho de que las comparativas proporcionales constituyen una

estructura que, desde el punto de vista lógico-semántico, pone en relación una causa con su consecuencia, en CSA se denomina a sus funciones constitutivas 1.º TÉRMINO (la causa lógica, *cuanto más aumentaban los planes de la fiesta*) y 2.º TÉRMINO (la consecuencia lógica, *más ideas de delirio se le ocurrían para hacerla más grande*). Se observa, además, que *cuanto más* y *más* ejercen funciones sintácticas (COMPLEMENTO CIRCUNSTANCIAL y DETERMINANTE en el segmento *más ideas de delirio*) en la *estructura comparativa*.

Las comparativas proporcionales son también reconocidas como tales en el corpus *UAM Spanish Treebank*. En *Lo mejor es buscar un término medio, pero siempre teniendo en cuenta que cuanto mayor sea el peso del usuario, más duro tendrá que ser el colchón* se reconoce la estructura comparativa que hay en *cuanto mayor sea el peso del usuario, más duro tendrá que ser el colchón*.

El corpus *AnCora* no muestra, por el contrario, un tratamiento homogéneo de las proporcionales, pues aunque en él se tiende a considerarlas como coordinadas (*Cuanto más atención se presta al texto, más se acentúan en el personaje exento los rasgos menos inocentes de un prototipo sobrecargado*), también es posible encontrarlas analizadas como condicionales (*Cuanto más frío y seco sea el aire, más rápidamente se formarán los copos*) o adverbiales a secas, es decir, sin el valor atributivo de comparativa. Así en

El presidente del PNV, Xabier Arzalluz, afirmó hoy que ETA «sabe muy bien que nunca ganará la guerra» y que no entiende que «cuanto más use las armas, más obstaculizará la política nacionalista unida y retrasará la fecha de nuestra independencia» no se reconoce una comparativa en cuanto más use las armas, más obstaculizará la política nacionalista unida y retrasará la fecha de nuestra independencia.

No hay, por último, ejemplos de comparativas proporcionales en el corpus *IULA Spanish LSP Treebank*.

4.4.

Comparativas superlativas

Las comparativas superlativas (*el/la/los/las/lo más... de/que*) son reconocidas como tales en CSA. Así en el ejemplo

- 6 la parranda de mayor escándalo que se había visto jamás en el pueblo **TÉRMINO Estructura comparativa**
- 7 la parranda de mayor escándalo **TÉRMINO PONDERADO Frase nominal**
- 8 la **DETERMINANTE Determinante** el
- 8 parranda de mayor escándalo **NOMINAL Frase sustantiva**
- 9 parranda **NÚCLEO Sustantivo** parranda
- 9 de mayor escándalo **MODIFICADOR Frase preposicional**
- 10 de **DIRECTOR Preposición** de
- 10 mayor escándalo **TÉRMINO Frase sustantiva**
- 11 mayor **MODIFICADOR Adjetivo** mayor
- 11 escándalo **NÚCLEO Sustantivo** escándalo
- 7 que se había visto jamás en el pueblo **TÉRMINO CLASE Cláusula**

Figura 4. Análisis en CSA de la *parranda de mayor escándalo que se había visto jamás en el pueblo*

vemos que las superlativas se tratan como un subtipo de *Estructura comparativa* desempeñada por las funciones **TÉRMINO PONDERADO** (*la parranda de mayor escándalo*) y **TÉRMINO CLASE** (*que se había visto jamás en el pueblo*).

Las comparativas superlativas, en contra de lo que sucedía en el caso de las proporcionales, no son reconocidas como tales en el corpus *UAM Spanish Treebank*: *¿Qué es lo más importante que ha aprendido mientras construía su imperio de Internet?*, por ejemplo, no contiene una estructura comparativa en *lo más importante que ha aprendido mientras construía su imperio de Internet*.

Las comparativas superlativas no son reconocidas como tales tampoco en el corpus *AnCora*, que no anota este tipo de construcciones de modo diferente a como anota cualquier frase nominal compleja. Así en

Los espejismos, según el diccionario, no son más que meras ilusiones ópticas, pero yo, que nunca he sido personaje de novela ni compañero de aventuras de Lawrence de Arabia, lo más parecido a un espejismo que he visto son esos bichitos blancos con forma de paramécio que culebreaan sobre el forro oscuro de los

párpados no se identifica una estructura comparativa en lo más parecido a un espejismo que he visto.

Las comparativas superlativas no son, por fin, anotadas como tales en el corpus *IULA Spanish LSP Treebank*. Así *Un virus es la más pequeña estructura biológica que contiene toda la información necesaria para su propia reproducción* no contiene una estructura comparativa en *la más pequeña estructura biológica que contiene toda la información necesaria para su propia reproducción*.

5. Conclusiones

A partir fundamentalmente del modelo descriptivo recogido en *NGLE* de la *RAE*, en el corpus *CSA* se ha formalizado la descripción de las comparativas de grado, tanto de las de 2.º término introducido por *que/como*, como de las de 2.º término introducido por *de*, de las comparativas proporcionales y de las comparativas superlativas. El reconocimiento de todos estos tipos de comparativas no es extensivo a otros corpus del español.

Se ha revisado en especial lo que ocurre con las comparativas de grado: siendo reconocidas como tales en *CSA*, su descripción se ha formalizado identificando como estructura comparativa el grupo cuantificativo y el segundo término, sin contar con el primero; entendiendo además que el segundo término se relaciona con el grupo cuantificativo y no con el cuantificador, y que puede tener estructura de construcción verbal. De la misma manera que con los tipos de comparativas en general, esta visión de las comparativas de grado no se extiende a todos los corpus que las reconocen.

Toda esta variación, sostenemos, procede del hecho, legítimo, de que distintos corpus pueden proponerse alcanzar un nivel de detalle u otro, pero también de que, incluso cuando se proponen llegar al mismo nivel de análisis desde un posicionamiento teórico-descriptivo similar, distintos corpus salvan la distancia entre una descripción comprensiva y una descripción formalizada

cada uno a su manera. Quizá deberíamos reflexionar, sobre todo, las instituciones con más recursos para ello, sobre lo mucho que lo señalado perjudica al intercambio y a la reutilización de unos recursos tan costosos de producir como son los corpus anotados, y proponernos alcanzar un mayor grado de formalización, al menos para propósitos como la elaboración de corpus, en las descripciones que desarrollamos.

6. Agradecimientos

Trabajo financiado por la Xunta de Galicia, proyecto DRASAE (10PXIB204269PR). Esta investigación ha podido también beneficiarse de la financiación FEDER/Ministerio de Ciencia, Innovación y Universidades – Agencia Estatal de Investigación/ESLORA+ (FFI2017-86379-P). M.^a Paula Santalla forma asimismo parte del grupo Gramática del Español de la Universidad de Santiago de Compostela, beneficiario de una ayuda para «Consolidación e estruturación de Grupos con Potencial de Creceemento 2017» de la Consellería de Cultura, Educación e Ordenación Universitaria de la Xunta de Galicia (ED431B 2017/39).

Referencias bibliográficas

- Brucart, J. M.^a (2003), “Adición, sustracción y comparación: un análisis composicional de las construcciones aditivo-sustractivas del español”, en F. Sánchez (ed.), *Actas del XXIII Congreso Internacional de Lingüística y Filología Románica*, Tübingen: Niemeyer, 11–60.
- Gutiérrez, S. (1994), *Estructuras comparativas*, Madrid: Arco/Libros.
- Marimon, M., Fisas, B., Bel, N., Arias, B., Vázquez, S., Vivaldi, J., Torner, S., Villegas, M., Lorente, M. (2012), “The IULA Treebank”, en *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC’12)*, Estambul: European Language Resources Association (ELRA), 1920–1926.

- Martí, M.^a A., Taulé, M., Bertran, M., Màrquez, L. (2008), *AnCora: A Multilingual and Multi-level Annotated Corpus*, [en línea] <https://www.researchgate.net/publication/228349954_AnCora_Multilingual_and_multilevel_annotated_corpora/download>, [fecha de consulta: 19/11/2018].
- Moreno, A., López, S., Sánchez, F. (1999), *Spanish Tree Bank: Specifications, Version 5* (30 April 1999), [en línea] <<http://www.lllf.uam.es/ESP/Treebank.html>>, [fecha de consulta: 19/11/2018].
- Real Academia Española (2009), *Nueva gramática de la lengua española. Morfología y sintaxis*, Madrid: Espasa Calpe.
- Real Academia Española (2010), *Nueva gramática de la lengua española. Manual*, Madrid: Espasa Calpe.
- Sáez, L. Á. (1999), “Los cuantificadores: las construcciones comparativas y superlativas”, en I. Bosque y V. Demonte (eds.), *Gramática descriptiva de la lengua española*, Madrid: Espasa Calpe, 1129–1188.
- Sáez, L. Á., Sánchez, C. (eds.) (2014), *Las construcciones comparativas*, Madrid: Visor.
- Santalla, M.^a P. (2012), “La elaboración de un sistema de anotación sintáctica. Algunos problemas relacionados con el tratamiento de la coordinación que implica a verbos”, en T. Jiménez Juliá, B. López Meirama, V. Vázquez Rozas y A. Veiga (eds.), *Cum corde et in nova grammatica. Estudios ofrecidos a Guillermo Rojo*, Santiago de Compostela: Universidade de Santiago de Compostela, 771–790.
- Taulé, M., Martí, M.^a A., Recasens, M. (2008), “AnCora: Multi-level Annotated Corpora for Catalan and Spanish”, en N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odjik, S. Piperidis y D. Tapias (eds.), *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, 96–101.