

CONSIDERACIONES SOBRE LA NATURALEZA DE LOS NÚCLEOS SINTÁCTICOS. HACIA UNA REPRESENTACIÓN SINTÁCTICA DE DEPENDENCIAS

MARINA LLOBERES SALVATELLA
GRIAL, Universitat Oberta de Catalunya
marina.lloberes@ub.edu

IRENE CASTELLÓN MASSALLES
GRIAL, Universitat de Barcelona
icastellon@ub.edu

RESUMEN

En el análisis sintáctico automático, la definición de criterios lingüísticos para gramáticas basadas en conocimiento lingüístico permite de desarrollar recursos coherentes y consistentes. La construcción de EsTxala y CaTxala, dos gramáticas de dependencias del español y del catalán para FreeLing (un entorno de herramientas de Procesamiento del Lenguaje Natural), se ha llevado a cabo según el diseño previo de un repertorio de criterios ecléctico y crítico en relación con algunos de los formalismos lingüísticos implementados en el análisis automático del lenguaje, la Gramática de Dependencias y la Gramática Generativa. El objetivo de dicho repertorio es facilitar la coherencia y la consistencia de la representación sintáctica en el desarrollo de gramáticas para el análisis sintáctico automático.

PALABRAS CLAVE: análisis sintáctico automático, criterios sintácticos, Gramática de Dependencias, Gramática Generativa, español, catalán.

CONSIDERATIONS ABOUT THE NATURE OF SYNTACTIC NUCLEI. TOWARDS A SYNTACTIC REPRESENTATION OF DEPENDENCIES

ABSTRACT

The task about defining linguistic criteria for parsing linguistic-based grammars allows to build coherent resources. The development of EsTxala and CaTxala, Spanish and Catalan dependency grammars for FreeLing environment (a set of Natural Language Processing tools), was carried out regarding a set of linguistic criteria previously designed, which was developed like an eclectic and critic resource by means of the main linguistic formalisms implemented in parsing, the Dependency Grammar and the Generative Grammar. The main aim of this repertoire is to facilitate the coherence and the consistency of syntactic analysis into the development task of parsing grammars.

KEYWORDS: parsing, syntactic criteria, Dependency Grammar, Generative Grammar, Spanish, Catalan.

1. INTRODUCCIÓN

El análisis sintáctico automático es una tarea del Procesamiento del Lenguaje Natural (PLN) necesaria en los procesos de comprensión del lenguaje humano. Muchas aplicaciones de PLN (traducción automática, extracción de información, anotación automática de roles semánticos, etc.) requieren de una cierta profundidad en los análisis sintácticos para poder obtener una buena representación semántica sobre la que aplicar procesos posteriores. En consecuencia, durante los últimos años, este ámbito del PLN ha hecho grandes avances.

Dada la dificultad de asignar el árbol sintáctico adecuado a una oración, en los últimos años se han propuesto soluciones diversas a esta tarea. Una de las soluciones adoptadas más consensuadas por la comunidad investigadora ha sido dividir el proceso de análisis en dos etapas. En primer lugar, se aplica un analizador parcial que agrupa unidades léxicas de forma unívoca creando constituyentes de primer nivel, en algunas ocasiones constituyentes no completos. Este proceso es lo que llamamos análisis sintáctico parcial (Schmid 1994; Castellón et al. 1998). En segundo lugar, estos análisis parciales son conectados mediante un análisis completo que, en el área de PLN, denominamos análisis profundo. La primera etapa, el análisis parcial, parece que ha alcanzado índices de éxito suficientes, por lo que actualmente los esfuerzos están puestos en la segunda etapa, el análisis profundo.

Este segundo nivel se ha afrontado de diferentes maneras. Existen multitud de sistemas de análisis basados en conocimiento estadístico, sistemas que requieren de un corpus anotado manualmente para poder adquirir modelos de lenguaje para posteriormente aplicarlos a la resolución de las dependencias sintácticas (Magerman 1995; Collins 2000; Yamada y Matsumoto 2003; Nivre 2006; Koo et al. 2008). Aunque actualmente este tipo de sistemas es el mejor en la evaluación estándar, algunos de los problemas que presentan estos sistemas son la necesidad y dependencia de grandes corpus anotados manualmente en este nivel para poder aplicar mecanismos de aprendizaje y la poca transportabilidad, es decir, aplicar un modelo del lenguaje adquirido sobre otros corpus no muy similares al inicial produce porcentajes de error más altos (Comelles et al. 2010). Otra solución consiste en el desarrollo de gramáticas basadas en conocimiento lingüístico. Se trata de definir reglas sintácticas manualmente, basadas en corpus y en gramáticas descriptivas de las lenguas, para luego aplicarlas a corpus generales de la lengua.

En esta última línea se sitúa el trabajo que presentamos, el desarrollo de gramáticas de las lenguas española y catalana (Lloberes et al. 2010): EsTxala y

CaTxala.¹ Estas gramáticas, basadas en el formalismo de dependencias (Tèsnière 1959; Mel'čuk 1988), han sido diseñadas para el análisis sintáctico automático y, en concreto, en el entorno de la librería de herramientas de PLN FreeLing (Padró et al. 2010).²

En la actualidad, existen diversas gramáticas del inglés, como MaltParser (Nivre 2006), Minipar (Lin 1998), Connexor (Tapanainen y Järvinen 1998) o Link Grammar (Sleator y Temperley 1991). Sin embargo, en otras lenguas el desarrollo de este tipo de recursos es menor. En español, por ejemplo, existen algunas gramáticas como HISPAL (Bick 2006), Slot Unification Grammar (Ferrández et al. 2000) o Spanish Resource Grammar en HPSG (Marimón 2010) y, en el caso del catalán, existen gramáticas superficiales o parciales (Castellón et al. 1998) y, en algunos casos, con algún grado más de profundidad (Alsina et al. 2002).

EsTxala y CaTxala son recursos libres que proporcionan análisis sintácticos profundos y robustos. Ambas siguen el formalismo propuesto por el analizador de dependencias de FreeLing, TXALA (Atserias et al. 2005), que se basa en la escritura de reglas heurísticas y éstas, juntamente con el acceso a una serie de recursos lingüísticos externos, producen una estructura de análisis basado en la gramática de dependencias.

La gramática de dependencias tiene como origen los postulados del formalismo de dependencias propuesto en *Elements de syntaxe structurale* (Tèsnière 1959). Este formalismo (Tèsnière 1959; Mel'čuk 1988) concibe la oración como una red de conexiones entre las unidades léxicas basada en las relaciones que se establecen entre ellas. A diferencia del análisis de constituyentes (Chomsky 1957), donde los nodos terminales se proyectan en unidades más complejas o constituyentes para expresar dichas relaciones, las relaciones de dependencia se establecen entre los mismos nodos terminales, es decir, las palabras.

Así, todas las formas (palabras) o bien dependen de otras o bien otras dependen de ellas, aplicando el concepto de núcleo estructural. Además, en este formalismo, las relaciones de dependencia se etiquetan con las funciones sintácticas. Aunque el formalismo de dependencias clásico determina que para cada nodo dependiente existe un único núcleo sintáctico, hay autores (De Marneffe et al. 2006) que proponen un análisis en que los nodos dependientes pueden aceptar dos núcleos sintácticos.

¹ ExTxala y CaTxala se han desarrollado en el marco de los proyectos KNOW (Ministerio de Educación y Ciencia, TIN2006-1549-C03-02) y KNOW2 (Ministerio de Ciencia e Innovación, TIN2009-14715-C04-03, TIN2009-14715-C04-04), y también han sido utilizadas en los proyectos OpenTrad y EuroOpenTrad (Ministerio de Industria, Turismo y Comercio, Programa PROFIT, FIT-350401-2006-5), dos proyectos que tienen como objetivo desarrollar traductores basados en la transferencia para las lenguas oficiales del Estado Español (español, catalán, gallego y vasco) y para el inglés.

² <http://nlp.lsi.upc.edu/freeling/>

La representación propuesta por el formalismo de dependencias, aunque se sitúa en la sintaxis, es una representación muy próxima a la estructura semántica del predicado (Mel'čuk 1988). Es por ello por lo que el análisis de dependencias es el formalismo idóneo para hacer un análisis profundo y completo cercano a la representación de su significado. Esto ha producido que en las últimas décadas, en el área del PLN, haya crecido el interés en el desarrollo de sistemas de análisis sintáctico automático basado en las dependencias (Tapanainen y Järvinen 1998; Collins 2000; De Marneffe et al. 2006; Nivre 2006) con el objetivo de mejorar el nivel sintáctico en PLN. En español y en catalán existen muy pocos analizadores basados en este formalismo, como DILUCT (Calvo y Gelbukh 2006), DepPattern (Gamallo y González 2009), MaltParser (Nivre 2006) y Connexor (Tapanainen y Järvinen 1998).

Para la creación de una gramática computacional es fundamental disponer de un corpus representativo de la lengua para poder disponer de datos empíricos, ya que las muestras de lenguaje real contienen muchos usos no siempre tratados en las gramáticas descriptivas. Además, otro de los aspectos que creemos fundamental es la especificación de los criterios desarrollados en la elaboración de las reglas. Los criterios deben especificar tanto las condiciones de agrupación de constituyentes como la determinación del núcleo de los grupos formados. Como hemos dicho, en general asumimos la teoría expuesta en *Elements de syntaxe structurale* (Tesnière 1959) en el marco de la Sintaxis Estructural. Sin embargo, en algunas ocasiones, tomamos soluciones más eclécticas, ello es debido a la priorización de la proximidad de la representación semántica en nuestra propuesta. No pretendemos la demostración de la viabilidad de una teoría, sino más bien la construcción de un sistema que funcione y sirva de puente entre el análisis morfológico de las oraciones y su representación semántica.

En esta línea, en este artículo vamos a centrarnos en esta tarea: la elaboración de un repertorio de criterios lingüísticos para nuestras gramáticas que resuelva los puntos críticos en cuanto a la definición de núcleos sintácticos

La especificación de cómo analizar las diferentes estructuras ha implicado la creación en paralelo de un repertorio de estructuras que finalmente proponemos como corpus de evaluación cualitativa para cualquier gramática de las lenguas tratadas. Así, en este artículo presentamos los puntos críticos de los criterios elaborados, ilustrándolos con muestras de dicho corpus de evaluación.

La estructura del artículo será la siguiente. En §2, trataremos la naturaleza de los núcleos sintácticos desde dos perspectivas, el análisis de constituyentes a través de la propuesta *Goverment and Binding* (Chomsky 1981) y el análisis de dependencias, que tiene como origen los postulados propuestos por Tesnière (1959) y Mel'čuk (1988) dentro de la Sintaxis Estructural, para en último lugar explicar el punto de vista adoptado en las gramáticas de Txala. En las secciones posteriores nos centraremos en los criterios de algunos aspectos en los cuales no hay acuerdo entre estos formalismos lingüísticos (§3): el papel de la preposición

(§3.1), las oraciones subordinadas (§3.2), las estructuras comparativas (§3.3) y la representación de las oraciones coordinadas (§3.4). Por último, presentaremos las conclusiones del trabajo (§4).

2. SOBRE LA NATURALEZA DE LOS NÚCLEOS SINTÁCTICOS

En el análisis sintáctico automático, se han implementado diferentes formalismos gramaticales (*Phrase Structure Grammar*, *Dependency Grammar*, *Head-Driven Phrase Structure Grammar*, *Link Grammar*, etc.). No obstante, entre estos formalismos lingüísticos, los dos principales a partir de los cuales se han desarrollado muchos de los analizadores sintácticos automáticos son: *Phrase Structure Grammar* (Chomsky 1957) y *Dependency Grammar* (Mel'čuk 1988). El formalismo *Phrase Structure Grammar*, que llamamos análisis de constituyentes, se basa en la Gramática Generativa (GG). Por otro lado, a partir del formalismo *Dependency Grammar* (Mel'čuk 1988), que traducimos como Gramática de Dependencias (GD), se ha desarrollado el análisis de dependencias, de base estructuralista (Tesnière 1959).

Ambos formalismos parten de la idea de que las unidades léxicas contienen información sintáctica y semántica o, en términos más específicos, subespecificaciones sintácticas y semánticas que determinan la configuración de la oración. No obstante, como se observará, la GG y la GD difieren en la noción de categoría léxica, cosa que conlleva soluciones diferentes en cuanto a las categorías léxicas que pueden funcionar como núcleo sintáctico dentro de la estructura sintáctica.

2.1. La Gramática Generativa

El análisis de constituyentes parte de la base de que el lenguaje es una estructura abstracta de signos en forma de jerarquía. A partir de las estructuras más simples, que se corresponden a las unidades del lenguaje, se forman estructuras más complejas, que tienen como último término la unidad superior de la jerarquía que reúne todos los signos de la estructura, la oración.

El hecho de que las estructuras más simples permitan crear estructuras más complejas es debido a la combinación de las unidades atómicas del lenguaje. Puesto que el lenguaje se presenta de manera inherente en un orden lineal, las operaciones de combinación no se llevan a cabo aleatoriamente, sino que están motivadas por esa linealidad. Estas operaciones son en su naturaleza reglas de reescritura que definen la estructura sintáctica (1), conocidas también como reglas de estructura sintagmática.

(1)

- i. $p \rightarrow p_1 p_2 p_3 \dots p_n$
- ii. $u \rightarrow u_1 u_2 u_3 \dots u_n$
- iii. $w \rightarrow p u$

Estas reglas establecen que existe un conjunto de categorías básicas (nombre, verbo, adjetivo, etc.), las categorías terminales, que definen el lexicon (2ii). Los terminales, pues, son la unidad sintáctica básica a partir de la cual se construyen las relaciones en el marco de la oración mediante expansiones o, en términos generativistas, proyecciones (2iii). Estas proyecciones tienen como finalidad agrupar los terminales en unidades sintácticas no terminales más complejas, los constituyentes, que ponen de manifiesto la estructura sintáctica de la oración. Los elementos que determinan la organización de la oración son los núcleos sintácticos. Por esa razón, las expansiones siguen el camino de cada núcleo y especifican el tipo de constituyente (Fig. 1).

(2)

- i. La noia menja
'La chica come'
- ii. Terminales
DT = { La }
N = { noia }
V = { menja }
- iii. Proyecciones
SN \rightarrow DT N
SV \rightarrow V
O \rightarrow SN SV

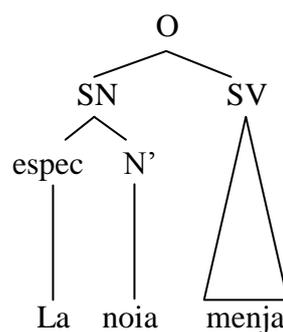


Fig. 1. Representación de la estructura sintáctica según la GG del ejemplo (2i) 'La noia menja' ('La chica come')

En la oración (2i) representada sintácticamente en Fig. 1, las tres categorías expresadas (DT, N, V) se proyectan en unidades más abstractas mediante las reglas de estructura sintagmática (2iii). De modo que el determinante 'la' y el nombre 'noia' se proyectan como *espec* y *N'* respectivamente, y, a su vez, ambas proyecciones se expanden una vez más debido a la existencia de una regla que

define la anidación de *espec* y *N'* como constituyentes del sintagma nominal (SN → DT N). Por otro lado, el verbo 'menja' se proyecta como *V*, *V'*, etc., y, en último término, como *SV*. Finalmente, la regla que define la formación de la oración (O → SN SV) permite anidar el sintagma nominal *SN*, expresado léxicamente mediante 'la noia', y el sintagma verbal *SV*, representado léxicamente por 'menja'.

2.2. La Gramática de Dependencias

En el formalismo de dependencias (Tèsnière, 1959; Mel'čuk, 1988), se considera que las unidades léxicas están conectadas porque entre ellas se establecen relaciones. En otras palabras, cada unidad léxica de una oración está relacionada sintácticamente con otra unidad léxica y no hay ninguna unidad léxica que quede fuera del alcance de la oración. Por lo tanto, la estructura sintáctica es un grafo de conexiones que pone de manifiesto las relaciones entre las unidades léxicas.

Mientras que en la GG las unidades léxicas se proyectan en nodos abstractos para construir la estructura sintáctica, desde la perspectiva de la GD, las relaciones entre las unidades léxicas son directas, es decir, cada unidad léxica está gobernada directamente por otra unidad léxica (3ii). De modo que en una relación sintáctica la unidad léxica dominante es el núcleo sintáctico de la unidad léxica dependiente. Así pues, la estructura sintáctica es un grafo de conexiones directas y jerárquicas (Fig. 2).

- (3)
- i. La noia menja
 'La chica come'
 - ii. menja ← noia
 noia ← La

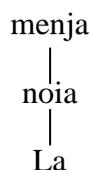


Fig. 2. Representación de la estructura sintáctica según la GD del ejemplo (3i) 'La noia menja' ('La chica come')

El hecho de que las unidades léxicas se conecten alrededor de una jerarquía permite que un núcleo sintáctico actúe, a su vez, como un nodo dependiente de otro nodo que se encuentra en un nivel superior de la jerarquía. Del mismo modo, un nodo dependiente puede ser el núcleo sintáctico de los nodos que dependen de éste. De todas formas, el nodo dependiente siempre está asociado a un solo núcleo sintáctico (Fig. 3) y, en cambio, un núcleo sintáctico puede dominar diversos nodos dependientes, a excepción de un caso. El núcleo sintáctico de la oración no

está gobernado por ningún otro núcleo sintáctico. Así pues, la estructura sintáctica es un grafo acíclico de conexiones directas y jerárquicas (Fig. 3).

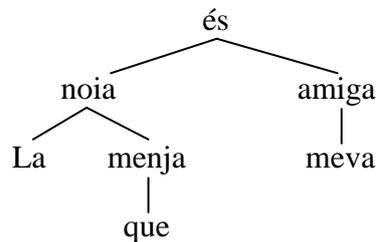


Fig. 3. Representación de la estructura sintáctica según la GD de la oración 'La noia que menja és amiga meva' ('La chica que come es amiga mía')

A pesar de que Tesnière (1959) y Mel'čuk (1988) consideran que no es posible que un nodo dependiente pueda estar gobernado por dos núcleos sintácticos, hay autores que defienden la duplicidad de núcleos sintácticos (De Marneffe et al. 2006) y, por lo tanto, consideran lícita la existencia tanto del análisis representado en Fig. 3 como del análisis ilustrado en Fig. 4.

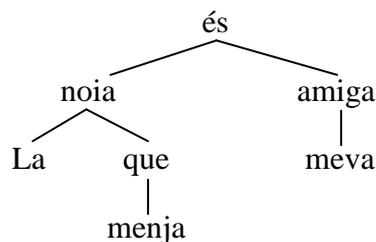


Fig. 4. Representación de la estructura sintáctica según la GD de la oración 'La noia que menja és amiga meva' ('La chica que come es amiga mía')

En cuanto a la determinación de las relaciones sintácticas, en ningún caso es aleatoria ya que está motivada por la semántica. Cada unidad léxica contiene subespecificaciones que distinguen las relaciones sintácticas entre las unidades léxicas (4).

(4)

- i. La profesora lleva a los alumnos al teatro
- ii. La profesora lleva a los alumnos lecturas nuevas

Por lo tanto, no es suficiente la estructuración de las unidades léxicas en el marco de la oración para explicar las relaciones de la oración. La naturaleza de las relaciones donde participan las unidades léxicas da sentido al conjunto de relaciones de la estructura sintáctica. Por ese motivo, en los ejemplos de (4), a pesar de que el verbo sea el mismo en ambos ejemplos ('llevar'), la relación establecida entre 'llevar' y 'a los alumnos' es diferente en cada caso.

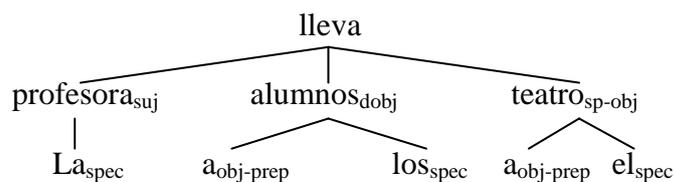


Fig. 5. Análisis de funciones sintácticas según la GD del ejemplo (4i) 'La profesora lleva a los alumnos al teatro'

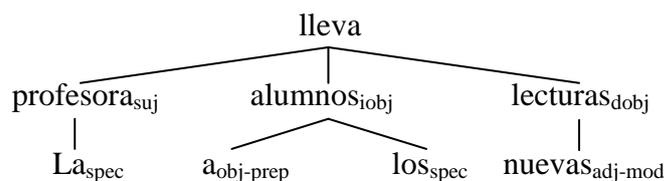


Fig. 6. Análisis de funciones sintácticas según la GD del ejemplo (4ii) 'La profesora lleva a los alumnos lecturas nuevas'

En (4i), aparece un objeto directo y, en (4ii), la relación establecida es la de objeto indirecto. De modo que las relaciones sintácticas aportan más información a la organización de las unidades léxicas dentro de la estructura sintáctica (Fig. 5, Fig. 6). En realidad, éstas actúan como un enlace entre la prosodia y la morfología, y el significado.

2.3. Las categorías léxicas y los núcleos sintácticos

Implicítamente, en los apartados anteriores (§2.1, §2.2), se ha dejado entrever que la estructuración de las unidades léxicas dentro de la jerarquía no ocurre de forma aleatoria, ya que los constituyentes que las representan son distintos (Hernanz y Brucart 1987:31).

Los constituyentes se clasifican en distintas categorías sintácticas y dicha clasificación está determinada por la naturaleza de su núcleo sintáctico (Hernanz 2002:1019). Por ejemplo, en el sintagma *la revista cultural*, se habla de sintagma nominal ya que el núcleo de este sintagma se corresponde a un nombre; a su vez, el sintagma *leen muchos libros* es categorizado como sintagma verbal porque el núcleo sintáctico está formado por un verbo.

El hecho de que se pueda distinguir, por ejemplo, un nombre de un verbo es debido al conjunto de propiedades que definen el repertorio de categorías léxicas. Estas categorías pueden ser flexionadas, forman repertorios abiertos, son semánticamente plenas y son categorías mayores (ya que pueden seleccionar complementos y son independientes morfológicamente).

Por otro lado, existe otro conjunto de categorías, las categorías gramaticales, que Trask (1992:122) define como:

gramatical category *n.* Any of various distinctions within which are expressed by variations in the form of lexical or phrasal constituents. [...]

Tal como describe Hernanz (2002:1021), estas categorías se diferencian de las categorías léxicas debido a que pertenecen a repertorios cerrados, generalmente no tienen independencia morfológica ni fonológica y son de carácter relacional (como consecuencia, no contribuyen de la misma manera a la interpretación semántica de la oración).

A pesar de que intuitivamente se puedan reconocer clases distintas de palabras, no hay acuerdo alguno sobre cuáles son las categorías que forman las clases de palabras (Hernanz y Brucart 1987:32). Las diferentes clasificaciones que se han propuesto coinciden a reconocer ciertas clases, como el nombre o el verbo, pero existen divergencias para admitir como clases de palabras el resto (como el adjetivo, la preposición, el adverbio, la conjunción).

Según la perspectiva de la GG, la naturaleza de la unidad léxica determina la categoría sintáctica del núcleo sintáctico (Chomsky 1981; Haegeman 1991:30). A diferencia de los otros formalismos gramaticales, la GG considera que todas las categorías léxicas son endocéntricas (es decir, categorías que contienen un núcleo y que son de igual distribución) y, como consecuencia, tienen la capacidad para actuar como núcleo sintáctico (Hernanz y Brucart 1987: 36).

En otro estadio diferente de la GG, se sitúa la GD (Tèsnière 1959; Mel'čuk 1988). Este formalismo reconoce la existencia de unidades léxicas (o 'palabras plenas' según la GD; Tesnière, 1959: 53), que están dotadas de función semántica porque por ellas mismas representan una idea. En contraposición, otro grupo de palabras forman la clase de unidades funcionales o, tal y como Tesnière (1959: 53) las denomina, 'palabras vacías'. Se trata de unidades sin carga semántica que actúan como meros mecanismos gramaticales o de relación de unidades léxicas (Tesnière 1959: 53). Según este punto de vista, las unidades léxicas son aquellas que pueden optar a funcionar como núcleos sintácticos y, en cambio, las unidades funcionales actúan alrededor de esos núcleos (por lo tanto, no ocupan posiciones nucleares en la estructura sintáctica).

3. UNA PROPUESTA DE REPRESENTACIÓN DE LOS NÚCLEOS SINTÁCTICOS

EsTxala y CaTxala (Lloberes et al. 2010) son dos gramáticas para el análisis sintáctico profundo del español y catalán que utilizan información lingüística como fuente de conocimiento y están basadas en la teoría de la GD (Tèsnière 1959; Mel'čuk 1988). De modo que las reglas que forman ambas gramáticas se fundamentan en los postulados del formalismo de dependencias.

No obstante, estas gramáticas son críticas en la consideración de la naturaleza de algunos núcleos sintácticos, ya que el objetivo es aportar análisis

sintácticos donde la representación de la estructura sintáctica sea próxima a estructura semántica del predicado (Mel'čuk 1988). Como consecuencia, proponen una alternativa de representación de algunas construcciones sintácticas. Esta propuesta combina criterios basados en la GD (Tèsnière 1959; Mel'čuk 1988) y la GG (Chomsky 1957). Por lo tanto, los criterios de representación aquí propuestos pretenden ser una aportación ecléctica y útil para otras gramáticas desarrolladas en el mismo ámbito del PLN.

A continuación, presentamos una revisión crítica de las estructuras sintácticas que comportan una representación sintáctica diferente en función de la teoría sintáctica, GG y GD. En concreto, se describen esas estructuras sintácticas y se exponen las razones que llevan a la GG y la GD a proponer análisis divergentes. Como consecuencia, el análisis de las propuestas teóricas permite optar por una de las soluciones y definir un criterio para EsTxala y CaTxala.

3.1. La preposición como unidad con significado pleno

La naturaleza sintáctica de la preposición ha sido protagonista de abundantes dialécticas teóricas. A lo largo de la historia, los estudios lingüísticos y filológicos se han centrado en dilucidar el comportamiento de la preposición respecto al resto de categorías léxicas (Fabra 1918; Badia i Margarit 1962; Alarcos 1994; Bonet y Solà 1986; Hernanz 2002). Mientras que tradicionalmente el nombre, el adjetivo y el verbo han sido consideradas categorías léxicas como tales, los teóricos discrepan a la hora de determinar si la preposición pertenece realmente o no a esas categorías.

La gramática tradicional (Fabra 1918; Badia i Margarit 1962; Alarcos 1994) propone una noción de la preposición, con frecuencia, vaga. A veces, incluso, algunos usos de la preposición se engloban en los usos de los adverbios o de expresiones adverbiales (Moreno Cabrera 1991:391). De modo que, desde un punto de vista tradicional, existe cierta confusión en la distinción entre preposición y adverbio (Tèsnière 1959:52; Bonet y Solà 1986:64). Prueba de ello es la clasificación de la preposición en usos transitivos (5i), donde la preposición no puede aparecer sin la presencia de un complemento, y en usos intransitivos (5ii), en los cuales el complemento de la preposición es opcional (Sancho Cremades 2002:1693) y que la gramática tradicional denomina adverbios utilizados como preposiciones (Bonet y Solà 1986:64).

(5)

- i. El pasillo conduce al comedor
- ii. Encontrarás las llaves { encima de la mesa / encima }

En cambio, estudios teóricos y descriptivos más recientes apuntan que la preposición, aunque mantiene un comportamiento distinto a las categorías léxicas

mayores, se pueden reconocer comportamientos afines a esas categorías (Bonet y Solà 1986; Hernanz y Brucart 1987; Hernanz 2002).

A diferencia de las categorías léxicas mayores, que, como se ha observado en §2.3, son variables, pertenecen a clases de palabras abiertas, su contenido es descriptivo, seleccionan complementos y son independientes morfológicamente (Hernanz 2002:1019), la preposición muestra un comportamiento distinto ya que es invariable, su significado es más abstracto que las categorías léxicas mayores, y no puede ocurrir sin un complemento (6).

(6)

- i. Laia escucha [_{SN} música]
- ii. Laia parece [_{SAdj} triste]
- iii. Laia [_{SV} ríe]
- iv. Laia nada [_{SAdv} bien]
- v. *Laia confía [_{SP} en]

De modo que el sintagma nominal, el sintagma adjetival, el sintagma verbal y el sintagma adverbial son, en realidad, sintagmas endocéntricos, es decir, sintagmas que se forman sobre la base de un núcleo (6i-iv). En contraposición, según la Sintaxis Estructural, el sintagma preposicional es de naturaleza exocéntrica ya que carece de núcleo como la oración.

No obstante, puesto que la GG considera que todas las proyecciones sintácticas son endocéntricas (Chomsky 1957; Bonet y Solà 1986; Hernanz y Brucart 1987), la preposición también presenta esa característica, pero a diferencia de las otras categorías léxicas, subcategoriza un complemento (Bonet y Solà 1986:64; Hernanz y Brucart 1987:36); de ahí que la oración de (6v) sea agramatical.

La hipótesis generativa de la endocentricidad de la preposición se confirma por el comportamiento paralelo de algunos verbos transitivos (7), que sólo aceptan la diátesis transitiva. El hecho de que (7ii) sea una estructura agramatical no quiere decir que el verbo sea una categoría exocéntrica. En tal caso, no se podría reconocer el verbo como categoría léxica (Hernanz y Brucart 1987:36). El verbo es una clase de palabra considerada categoría léxica, ya que el sintagma verbal siempre está dotado de núcleo. Entonces, el factor que determina la agramaticalidad de (7ii) reside en las propiedades sintáctico-semánticas de ese verbo: el verbo 'saber' prevé en su estructura un argumento que funciona como objeto directo. Por lo tanto, la ausencia de dicho argumento causa la agramaticalidad de la construcción.

(7)

- i. Supe que se había ido
- ii. *Supe

Debido a la caracterización de la preposición como categoría endocéntrica que propone la GG (Chomsky 1981:48; Bonet y Solà 1986:65; Hernanz y Brucart 1987:36), esta categoría admite proyecciones sintácticas de la misma forma que el resto de categorías léxicas, mediante lo cual la preposición tiene la capacidad de actuar como núcleo sintáctico (Fig. 7).

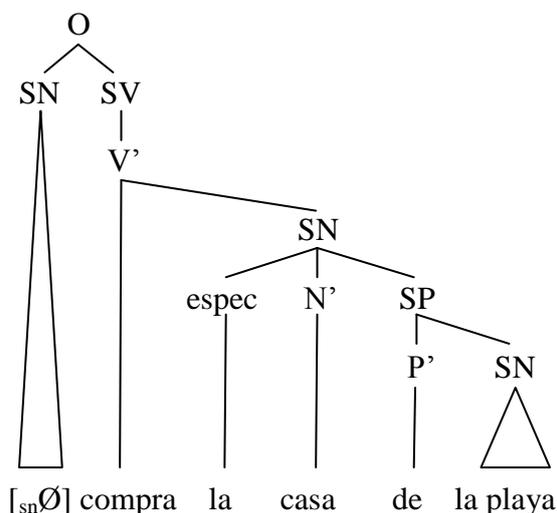


Fig. 7. Análisis del sintagma preposicional según la GG de la oración 'Compra la casa de la playa'

De manera similar a la gramática tradicional, la GG (Chomsky 1981; Bonet y Solà 1986; Hernanz y Brucart 1987) y la GD (Tesnière 1959; Mel'čuk 2003) coinciden en la consideración de la preposición como una pieza léxica relacional.

No obstante, la GD sitúa la preposición alejada de las categorías como el nombre, el adjetivo y el verbo (Tesnière 1959:53). Según la GD, la preposición es una categoría que no está cargada de ninguna función semántica y, por esa razón, se clasifica como una unidad funcional, aquellas unidades que únicamente aparecen en el discurso para indicar, precisar y transformar la categoría de las unidades léxicas y determinar las relaciones existentes entre ellas. Como consecuencia, la preposición no puede actuar como núcleo sintáctico (Fig. 8).

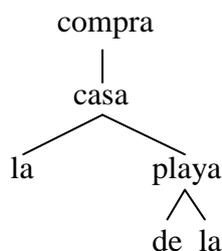


Fig. 8. Análisis del sintagma preposicional según la GD de la oración 'Compra la casa de la playa'

Puesto que EsTxala i CaTxala son gramáticas basadas en el formalismo de la GD, siguen sus postulados. No obstante, en cuanto a la naturaleza de la preposición, a diferencia de la GD (Tesnière 1959:53), se considera que es una pieza gramatical con función semántica, aunque mantiene diferencias con el resto de categorías léxicas mayores ya que funciona como un elemento relacional. La evidencia de la carga semántica de la preposición se halla en la posibilidad de distinguir preposiciones con significado propio y preposiciones con carácter semánticamente vacío (Hernanz y Brucart 1987:263). El primer tipo hace referencia a las preposiciones que introducen circunstanciales, un locativo en (8i), y el segundo se refiere a las preposiciones regidas por el verbo del predicado, un tema en (8ii).

(8)

- i. Los excursionistas andan por el campo
- ii. La empresa apuesta por inversiones en el extranjero

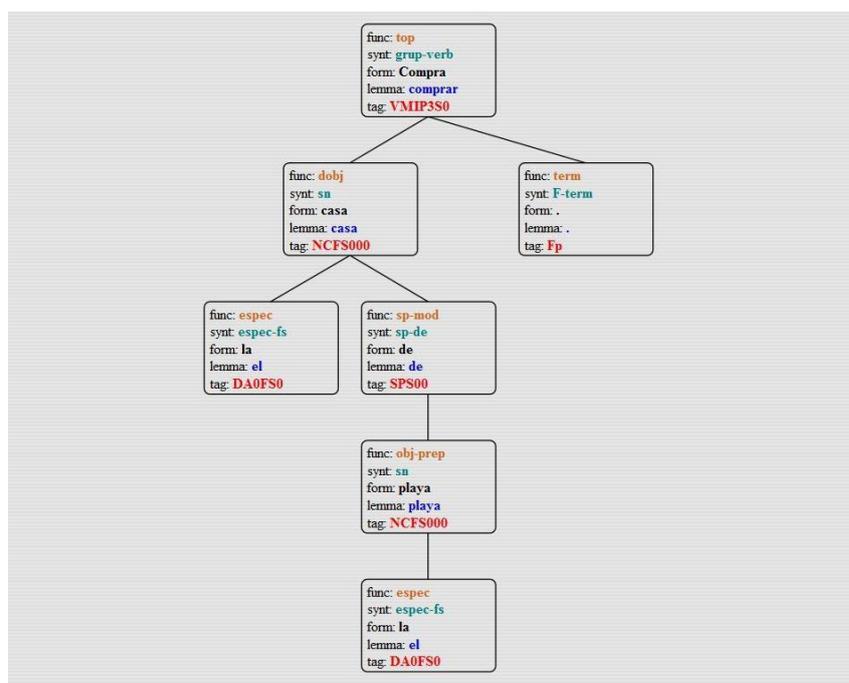


Fig. 9. Análisis del sintagma preposicional en EsTxala de la oración 'Compra la casa de la playa'

Igualmente, la posibilidad de que el verbo del predicado subcategorice un argumento introducido mediante una determinada preposición, como se observa en el ejemplo (8ii), es una muestra más de la relevancia de la preposición como categoría léxica. Por lo tanto, en las gramáticas de Txala, la preposición, junto con el resto de categorías léxicas, funciona como una unidad léxica con capacidad de actuar como núcleo en la estructura sintáctica (Fig. 9).

3.2. La variabilidad estructural de la subordinadas

Las subordinadas con verbo en forma personal o subordinadas en tiempo finito aparecen en la oración como argumento o adjunto de otro constituyente oracional modificándolo o complementándolo (Villalba 2002:2291). De todas formas, a diferencia de los constituyentes sintagmáticos, dicha relación se lleva a cabo mediante un operador gramatical (9) que expresa la relación entre ambas oraciones, principal y subordinada (9i', 9ii').

(9)

- i. Dijo que era verdad
- i'. [Dijo] *que* [era verdad]
- ii. Es van reunir on havien acordat
'Se reunieron donde habían acordado'
- ii'. [Es van reunir] *on* [havien acordat]

Sin embargo, de la observación de los ejemplos de (9), se perciben numerosas diferencias en el tipo de modificación o complementación. En (9i), la oración subordinada toma un valor sustantivo (Moreno Cabrera 1991:671; Delbecque y Lamiroy 1999:1967; Bonet 2002:2321), de ahí que se las conozca como subordinadas sustantivas. Como consecuencia, puede ser sustituida por un nombre sin que afecte a la gramaticalidad de la construcción (*Dijo la verdad*). En cambio, en (9ii), el valor de la subordinada es de naturaleza distinta, tiene un valor adverbial (Moreno Cabrera 1991:671; Villalba 2002:2251), y, por esa razón, se las denomina como subordinadas adverbiales. Por ello, un adverbio puede aparecer en su lugar (p.e. *Es van reunir allà* 'Se reunieron allí').

La modificación o complementación puede establecerse en otros niveles diferentes de la oración, por ejemplo a nivel de nombre (p.e. *No le gusta la idea de que vengan*, *La possibilitat que tanquin l'empresa és real* 'La posibilidad que cierren la empresa es real') o de adjetivo (p.e. *Está segura de que ha perdido las llaves*, *És una dona desitjosa que els seus fills tornin aviat* 'Es una mujer deseosa de que sus hijos vuelvan pronto'). En ese contexto, existen subordinadas en tiempo finito que típicamente modifican el nombre y que funcionan de manera similar a los adjetivos, las oraciones relativas (Moreno Cabrera 1991: 671; Brucart 1999: 397; Solà 2002: 2455; Villalba 2002: 2251). Este tipo de cláusulas tienen la particularidad de que el operador gramatical que las introduce es a la vez entendido como un nexos y como pronombre relativo anafórico (10).

(10)

- i. La chica que viene gruñendo es mi prima
- ii. La chica enfadada es mi prima

Por otro lado, en español y en catalán, es posible otro tipo de construcciones que están a medio camino entre las interrogativas indirectas y las relativas (Solà 2002: 2534). Aparecen como complemento o modificador del verbo principal de la oración y implícitamente expresan una interrogación, cosa que las asemeja al funcionamiento de las interrogativas indirectas (11i). No obstante, el operador gramatical que las introduce se corresponde al tipo de marcador típico de las relativas, pese a que en muchas ocasiones puede coincidir con el operador gramatical que introduce una interrogativa indirecta. Por esa razón, ese tipo de construcciones se conoce con el término pseudorelativas (11ii).

(11)

- i. Ha preguntado con quien habíamos hablado
- ii. Telefona a qui ha trobat al cinema
'Llama a quien ha encontrado en el cine'

La necesidad de la existencia de un operador gramatical que introduzca los diversos tipos de subordinadas finitas es una característica tanto del español como del catalán, a diferencia de otras lenguas como el inglés, que, con frecuencia, en determinados contextos permite la omisión de dicho marcador (p.e. *The book I was reading was quite interesting* 'El libro que estaba leyendo era bastante interesante'). No obstante, tal y como apunta Solà (2002:2509), la elisión del pronombre relativo había estado posible en otros estadios del catalán (12i). Por otro lado, son posibles estructuras con omisión del operador gramatical. En español, se pueden dar estos casos de omisión de conjunción (12ii) cuando la subordinada es expresada en modo subjuntivo y el sujeto de la oración principal no es expresado léxicamente (Delbecque y Lamiroy 1999:2026). Por el contrario, en catalán, parece ser una estructura muy restrictiva (12iii) o, más bien, extensamente condenada por la normativa (Bonet 2002:2348).

(12)

- i. Compta de la faina Ø ha feta Mestre Juan Estelrich [...] Dia 16 Janer 1799³
- ii. Se ruela no toquen los animales
- iii. *Li agraiim respecti el descans dels veïns
'Le agradecemos respete el descanso de los vecinos'

Evidentemente, los trabajos de lingüística descriptiva apuntan la existencia de otras construcciones relacionadas directamente o próximas a las subordinadas (Solà 1972; Bonet y Solà 1986; Bosque y Demonte 1999; Solà et al. 2002), como, por ejemplo, las interrogativas directas (*¿Quién ha dicho eso?*), las interrogativas indirectas (*No sap com tornarà a casa* 'No sabe cómo volverá a casa'), las subordinadas causales (*Va a perdre el avió perquè se ha dormido*), finales (*Esta*

³ Ejemplo del Arxiu del Regne de Mallorca citado por Solà (2002:2509).

exposición está pensada para que el visitante conozca nuevas realidades sociales), concesivas (*Anirà de vacances, encara que ha de treballar* 'Irà de vacaciones, aunque tiene que trabajar') y condicionales (*Puede saltarse clases siempre que traiga un justificante*). En nuestra propuesta, no hacemos un análisis detallado sobre la representación con fines computacionales de estas construcciones, ya que se asemejan a las construcciones que acabamos de describir. No obstante, nos referiremos a algunas de ellas a lo largo de esta sección.

Desde la perspectiva representacional, surgen divergencias a la hora de proponer una representación de las cinco construcciones descritas previamente: subordinadas sustantivas (13i), subordinadas adverbiales (13ii), relativas (13iii), pseudorelativas (13iv) y subordinadas con elisión de marcador (13v); puesto que, en función de los postulados teóricos, los formalismos gramaticales y, en nuestro caso, la GG y la GD proponen distintas soluciones a la representación de esas construcciones. La discusión gira alrededor de qué categoría gramatical actúa como núcleo de la subordinada: la conjunción subordinante o el verbo subordinado.

(13)

- i. El gerente ha decidido que los sueldos se congelen
- ii. Ha arribat a casa quan el partit havia acabat
'Ha llegado a casa cuando el partido había terminado'
- iii. La película que vio ayer le trajo recuerdos de su pasado
- iv. Es preocupa per quanta gent hi haurà
'Se preocupa por cuanta gente habrá'
- v. Rogamos vigilen sus pertenencias

3.2.1. Las subordinadas sustantivas y las subordinadas adverbiales

Dada la naturaleza lexicalista de la GG (Chomsky 1957; Bonet y Solà 1986; Haegeman 1991), la categoría gramatical que define una construcción subordinada es el complementante. En otros términos, el núcleo sintáctico de las subordinadas se corresponde a la conjunción tal y como se propone en el siguiente esquema derivado de las reglas de estructura sintagmática (Bonet y Solà 1986: 21; Haegeman 1991: 107). A pesar de eso, en función de la clase de subordinada o, más concretamente, en función del complementante, el esquema puede variar como veremos.

(14)

- i. $O \rightarrow SN SV^4$
- ii. $O' \rightarrow COMP O^5$

El esquema (14) explica la formación de la estructura de las subordinadas sustantivas (Bonet y Solà 1986: 21; Haegeman 1991: 107). El complementante en esta clase de subordinadas aparece como un mero operador gramatical que vincula la subordinada sustantiva y la oración principal, e insiere la subordinada sustantiva dentro de la estructura de la oración principal. De modo que la conjunción tiene como única función de marcador de inicio o encabezamiento de cláusula, como se puede observar en el ejemplo (Fig. 10).

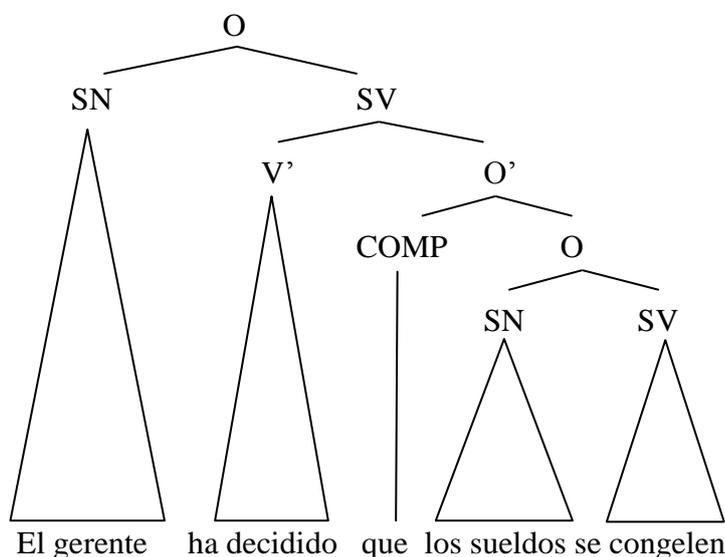


Fig. 10. Análisis de la subordinada sustantiva según la GG del ejemplo (13i),
'El gerente ha decidido que los sueldos se congelen'

⁴ En realidad, el esquema original incluye la categoría terminal inflexión (INFL), que reúne aspectos como tiempo y concordancia (INFL $\equiv [\pm$ Tiempo, (CONC)]), como nodo de la estructura ($O \rightarrow SN INFL SV$), ya que la GG considera que esta categoría es exocéntrica. No obstante, en este trabajo, no la representamos gráficamente porque presentamos una representación simplificada de la GG para facilitar la lectura de los análisis.

Por otro lado, cabe la posibilidad de otra interpretación en la regla de formación de la oración relacionada con la opcionalidad del sintagma nominal sujeto, ya que, con frecuencia, hay lenguas que permiten esta opcionalidad (*Plou* 'Llueve'). Como consecuencia, Chomsky (1981) propone una regla del tipo: $O \rightarrow (SN) INFL SV$.

⁵ En realidad, la regla ilustrada en (14) es incompleta, ya que la GG considera que la categoría COMP es extensible a cualquier tipo de cláusula, ya esté o no esté inserida en otra cláusula (c). De modo que el esquema (14) debería ser representado de la siguiente manera:

- $O \rightarrow SN SV$
- $O' \rightarrow COMP O$
- $O'' \rightarrow (TOP) O'$

Paralelamente, las subordinadas adverbiales presentan el mismo comportamiento que las subordinadas sustantivas, es decir, el complementante, materializado léxicamente por una conjunción, opera como un marcador de la subordinada (Bonet y Solà 1986:21; Haegeman 1991:107), con lo cual es el núcleo sintáctico de la cláusula (Fig. 11).

Por otro lado, basándonos en el formalismo de la GD (Tesnière 1959; Mel'čuk 2003), la conjunción forma parte del conjunto de unidades funcionales intranucleares (Tesnière 1959: 82), es decir, categorías gramaticales sin carga semántica que actúan como simples instrumentos gramaticales. Como consecuencia, esta categoría no puede funcionar en ningún caso como núcleo sintáctico de una subordinada sustantiva (Fig. 12) o de una subordinada adverbial (Fig. 13). En realidad, la categoría que introduce estas cláusulas subordinadas es el mismo verbo subordinado, ya que el verbo sí es una unidad léxica con carga semántica capaz de actuar como núcleo sintáctico.

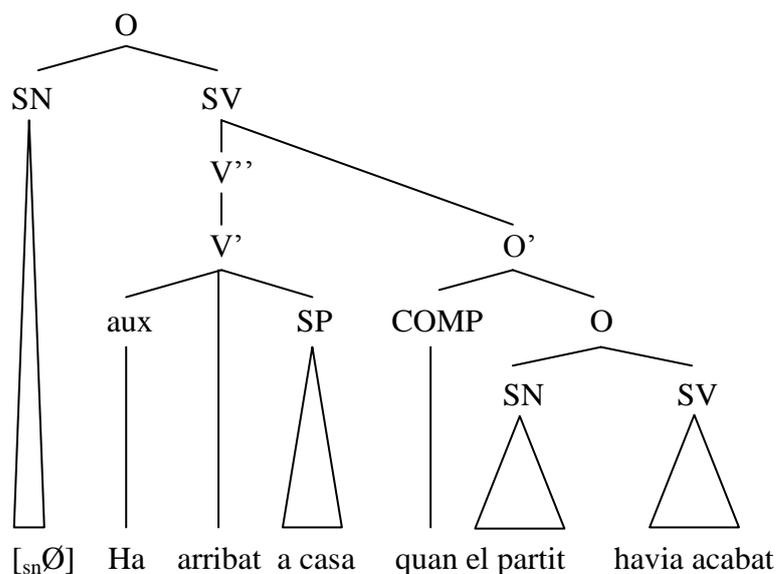


Fig. 11. Análisis de la subordinada adverbial según la GG del ejemplo (13ii), 'Ha arribat a casa quan el partit havia acabat' ('Ha llegado a casa cuando el partido había terminado')

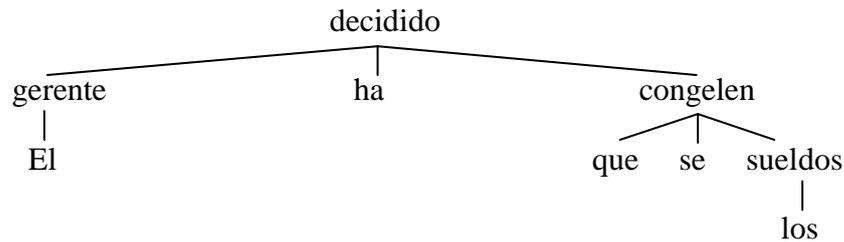


Fig. 12. Análisis de la subordinada sustantiva según la GD del ejemplo (13i),
'El gerente ha decidido que los sueldos se congelen'

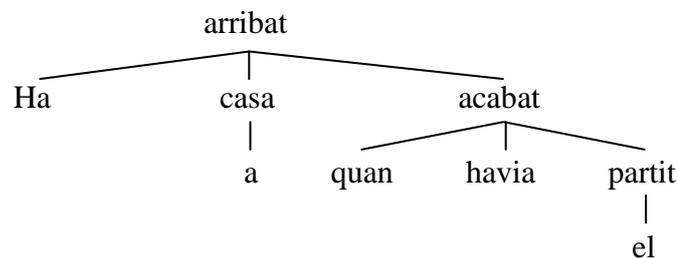


Fig. 13. Análisis de la subordinada adverbial según la GD del ejemplo (13ii),
'Ha arribat a casa quan el partit havia acabat' ('Ha llegado a casa cuando el partido había terminado)

A la hora de considerar cómo deben representarse ambas estructuras en las gramáticas de dependencias EsTxala y CaTxala, se parte de la idea de que la conjunción subordinante es un operador gramatical, de acuerdo con la GG (Chomsky 1957) y la GD (Tesnière 1959; Mel'čuk 2003). No obstante, a diferencia de la GD, en las gramáticas de Txala, se establece como la categoría sintáctica nuclear de la cláusula (Fig. 14, Fig. 15).

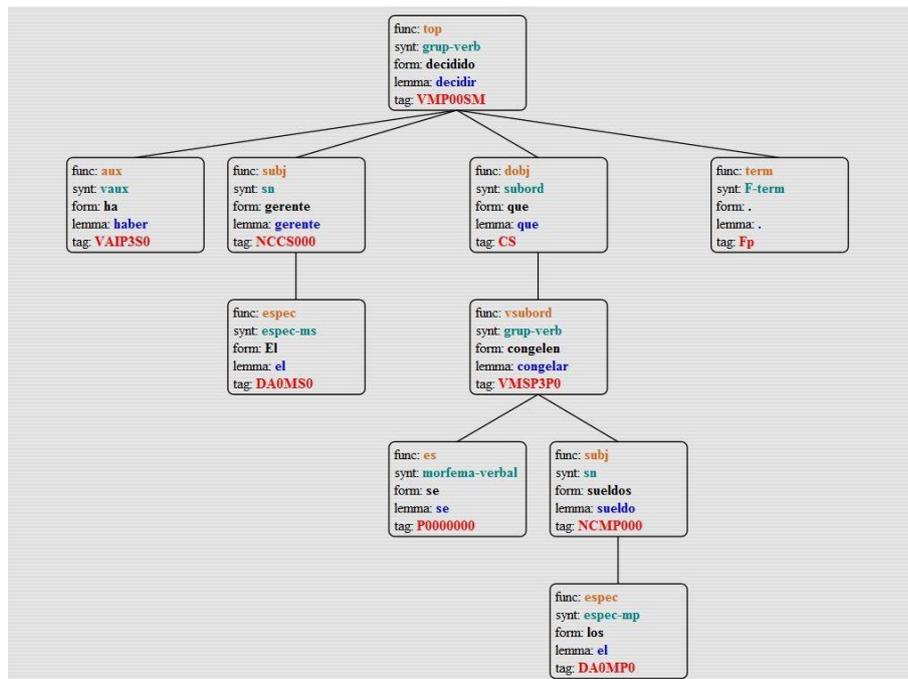


Fig. 14. Análisis de la subordinada sustantiva en EsTxala del ejemplo (13i), 'El gerente ha decidido que los sueldos se congelen'

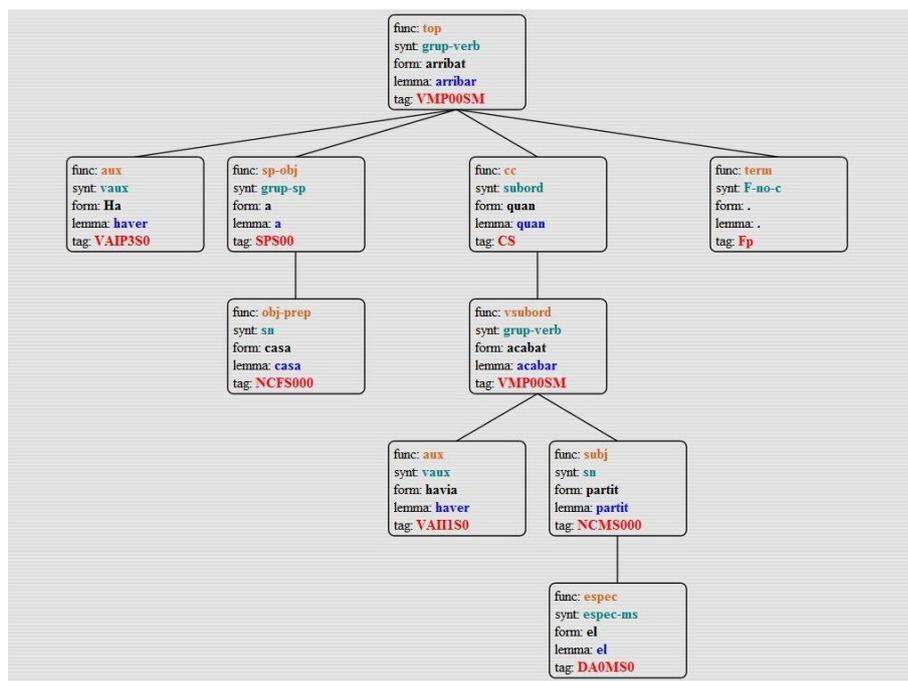


Fig. 15. Análisis de la subordinada adverbial en CaTxala del ejemplo (13ii), 'Ha arribat a casa quan el partit havia acabat' ('Ha llegado a casa cuando el partido había terminado')

El criterio establecido sigue el razonamiento generativista, ya que, precisamente, la conjunción subordinante es la categoría gramatical clave que define la construcción subordinada como tal (Moreno Cabrera 1991; Bonet y Solà 1986). Aunque cabe reconocer que este elemento nuclear mantiene numerosas

diferencias con otros núcleos sintácticos (nombre, adjetivo, verbo, adverbio), es el encargado de relacionar la oración principal y la oración subordinada. Por lo tanto, la estructura que proponemos para las subordinadas sustantivas y subordinadas adverbiales se puede resumir en el esquema (15).

$$(15) V_{\text{principal}} \leftarrow \text{CONJUNCIÓN} \leftarrow V_{\text{subordinado}}$$

Este esquema explica también la representación en las gramáticas de Txala de las interrogativas indirectas y las subordinadas causales, finales, concesivas y condicionales, porque, desde una perspectiva computacional, la estructura es idéntica a ambas subordinadas descritas: la anidación de nodos implicados se establece entre el verbo principal, la conjunción subordinante y el verbo subordinado.

3.2.2. Las oraciones relativas

Como hemos avanzado, el operador gramatical característico de las subordinadas relativas (16i) es sustancialmente diferente de las subordinadas que acabamos de observar. A rasgos generales, este marcador funciona como nexo de la misma manera como ocurre en las subordinadas sustantivas y en las subordinadas adverbiales. Pero, a su vez, mantiene una relación anafórica con el antecedente (Brucart 1999: 397; Solà 2002: 2459) y, en concreto, a modo de pronombre relativo (16ii). Como consecuencia de su naturaleza pronominal, el pronombre relativo se insiere dentro de la estructura argumental de la cláusula subordinada y lleva a cabo, además, una función sintáctica dentro de la subordinada (16iii).

(16)

- i. La película que vio ayer le trajo recuerdos de su pasado
- ii. La película ← que
- iii. Vio [OD la película] ayer

Debido a esta casuística, los formalismos gramaticales deberán tomar partido en esta cuestión y definirse para proponer una representación que diferencie las relativas de las subordinadas sustantivas y de las subordinadas adverbiales.

Desde la perspectiva generativista, se establece que el pronombre relativo introduce las relativas (Brucart 1999: 398; Solà 2002: 2459). De modo que este pronombre es el núcleo sintáctico de esta clase de subordinadas, cosa que se asemeja a las subordinadas que hemos observado (§3.2.1). No obstante, la GG reconoce que la representación de este marcador debe expresar la naturaleza pronominal de este constituyente.

Por ese motivo, se considera que los pronombres relativos son en realidad constituyentes-*qu* (Bonet y Solà 1986: 25; Haegeman 1991: 330),⁶ es decir, categorías que ocupan la posición de COMP y que están marcadas positivamente con el rasgo [+*qu*].⁷ Originalmente, estas categorías ocupan una posición en la estructura argumental del predicado verbal de la cláusula, pero, debido a una regla de movimiento de *qu* (Bonet y Solà 1986: 124; Haegeman 1991: 360), los constituyentes-*qu* se sitúan en la posición de COMP (Fig. 16).

Una concepción completamente diferente es la que propone la GD (Tesnière 1959; Mel'čuk 2003). De la misma forma que las conjunciones, como ya se ha observado, el pronombre relativo se incluye dentro del repertorio de unidades funcionales. Particularmente, según palabras de Tesnière (1959: 82), se trata de un traslativo, cuya función es "transformer la catégorie des mots pleins" (Tesnière 1959:82), es decir, operar sobre las unidades léxicas como categoría intranuclear (Tesnière 1959: 137). Por lo tanto, el pronombre relativo en ningún caso puede funcionar como núcleo sintáctico (Fig. 17).

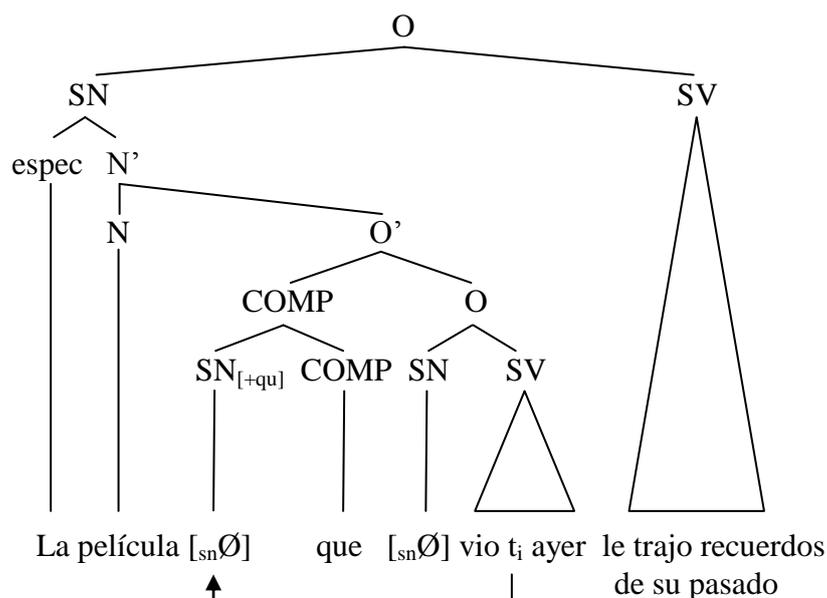


Fig. 16. Análisis de la subordinada relativa según la GG del ejemplo (13iii),
'La película que vio ayer le trajo recuerdos de su pasado'

⁶ Con el término 'constituyente-*qu*' nos referimos efectivamente al *wh-constituent* del inglés ya que las palabras que se corresponden a este constituyente empiezan *wh-* (Haegeman, 1991:330).

⁷ El rasgo [+*qu*] distingue, por ejemplo, los pronombres relativos de los elementos caracterizados por el rasgo [+QU] (Bonet y Solà, 1986:25). Este último rasgo es el encargado de identificar oraciones declarativas y no declarativas. En otros términos, el rasgo [+QU] se atribuye a la posición específica del complementante (COMP) y permite diferenciar los operadores gramaticales que introducen subordinadas sustantivas o de infinitivo (marcados negativamente, [-QU]) de los operadores propios de las interrogativas indirectas y directas (marcados positivamente, [+QU]).

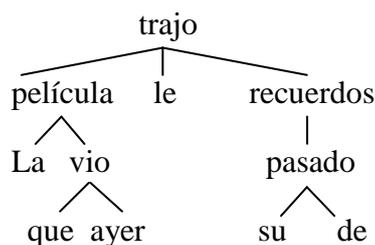


Fig. 17. Análisis de la subordinada relativa según la GD del ejemplo (13iii),
 ‘La película que vio ayer le trajo recuerdos de su pasado’

En ambas gramáticas de Txala, el pronombre relativo es tratado como un instrumento gramatical, es decir, como una categoría gramatical que relaciona cláusulas y marca la existencia de una subordinada, de acuerdo con las hipótesis de la GG (Bonet y Solà 1986; Haegeman 1991). No obstante, la representación que se hace del pronombre relativo difiere de la representación generativista y, de la misma forma que la GD (Tesnière 1959; Mel’čuk 2003), se considera que el núcleo sintáctico de las relativas es, en realidad, el verbo subordinado.

A pesar de optar por la misma representación que la GD, los motivos que llevan a dicho análisis son alejados de este formalismo gramatical. En EsTxala y CaTxala, el pronombre relativo no es entendido como una ‘palabra vacía’ (en términos de *Elements de syntaxe structurale*; Tesnière, 1959); aunque se reconoce también como marcador de cláusula. La razón de este criterio viene dada por la doble funcionalidad de este marcador en las relativas, como nexos y como pronombre relativo.

El pronombre relativo no es sólo un operador gramatical sino que también desarrolla una función dentro de la cláusula subordinada, que está supeditada a la información sintáctica y semántica contenida en la pieza verbal. Por lo tanto, debe ocupar una posición dentro de la cláusula, subordinada al verbo. En resumen, el esquema de las relativas es el siguiente (17).

$$17) V_{\text{principal}} \leftarrow V_{\text{subordinado}} \leftarrow PR_{\text{rel}}$$

3.2.3. Las pseudorelativas

La pseudorelativa es una construcción que aparentemente es ambigua sintácticamente con la interrogativa indirecta ya que ambas subordinadas pueden aparecer con los mismos marcadores (Solà 2002: 2534). Además, esta construcción ocupa la posición típica de las subordinadas sustantivas y de las subordinadas adverbiales (Moreno Cabrera 1991: 671; Delbecque y Lamiroy 1999: 1967; Bonet 2002: 2321; Villalba 2002: 2291).

A pesar de eso, no cabe duda de que la construcción pseudorelativa está introducida por un constituyente-*qu* marcado positivamente (es decir, [+*qu*]), según la GG (Bonet y Solà 1986: 25; Haegeman 1991: 330). Por lo tanto, se

diferencian de las interrogativas indirectas, cuyo operador gramatical está marcado con el rasgo [+QU] (Bonet y Solà 1986: 25). El análisis que ofrece la GG, pues, sitúa el constituyente-qu como el núcleo sintáctico de la subordinada. En este sentido, el análisis propuesto es paralelo a la estructura propia de las subordinadas relativas.

Por otro lado, la GD (Tesnière 1959; Mel'čuk 2003) considera que la estructura de las pseudorelativas es paralela a las relativas. Como consecuencia, el análisis que propone es el mismo que el análisis de las oraciones de relativo. Es decir, el operador gramatical de esta construcción es un pronombre relativo y, por lo tanto, se incluye como una categoría propia de las unidades funcionales y no tiene opción de actuar como núcleo de la cláusula ya que el verbo subordinado es el núcleo sintáctico que insiere la cláusula relativa en el marco de la oración principal.

En EsTxala y CaTxala, admitimos que esta construcción permite dos análisis, ya que nocionalmente la pseudorelativa es una relativa, pero, a su vez, ocupa la posición propia de un argumento del predicado verbal o de un adjunto. Por lo tanto, consideramos que esta construcción admite ambos análisis: el análisis previsto para las subordinadas sustantivas y subordinadas adverbiales (como consecuencia, el operador gramatical sería el núcleo sintáctico de la cláusula), y el análisis que proponemos para las relativas (es decir, la posición del núcleo sintáctico sería ocupada por el verbo de la subordinada relativa), como podemos ver en el esquema (17).

No obstante, puesto que la arquitectura del analizador Txala (Atserias et al. 2005) no permite proponer más de un análisis para cada secuencia lingüística de entrada, nos definimos por uno de los dos análisis posibles de la pseudorelativa. Como hemos expuesto, esta construcción es 'a grosso modo' una relativa en posición de argumento o adjunto. Optamos, pues, por el análisis que define esta construcción como una relativa (Tesnière 1959: 137). Por lo tanto, el núcleo sintáctico de la construcción es el verbo de la subordinada y el operador gramatical se anida como dependiente de este núcleo.

3.2.4. Las subordinadas con elisión de marcador

Como se ha descrito al inicio de esta sección, es posible la omisión del operador gramatical en determinados situaciones tanto en español como en catalán (Delbecque y Lamiroy 1999: 2026; Bonet 2002: 2348), aunque parece que el catalán es más restrictivo en esta cuestión o, como mínimo, esta elisión ha sido condenada por la normativa.

A pesar de esa controversia, ambos formalismos gramaticales aquí considerados, GG (Bonet y Solà 1986; Haegeman 1991) y GD (Tesnière 1959; Mel'čuk 2003), están de acuerdo con la representación de esta clase de elisión. La anidación se lleva a cabo entre el verbo principal de la oración y el verbo de la

subordinada, que actúa como núcleo sintáctico de la cláusula. De todas formas, la GG prevé una posición vacía de COMP ya que léxicamente no se expresa el marcador pero semánticamente está presente. Como consecuencia, en EsTxala y CaTxala, también se opta por esa representación.

Debido a la existencia de subordinadas con elisión del operador gramatical y la consiguiente representación $V_{\text{principal}} \leftarrow V_{\text{subordinado}}$, se podría contradecir el criterio de representación de las subordinadas sustantivas y subordinadas adverbiales, que establece que el núcleo sintáctico de la cláusula subordinada debe ser la conjunción. No obstante, creemos que es necesario distinguir estas subordinadas, donde el marcador es sólo un elemento relacional de cláusulas, de las relativas, donde el marcador desarrolla una función sintáctica dentro de la cláusula.

3.2.5. Resumen

En esta sección, se ha descrito y detallado el conjunto de subordinadas que se toman en consideración en las gramáticas de dependencias del español y del catalán para fines de procesamiento automático del lenguaje (Lloberes et al. 2010). Con el objetivo de establecer los criterios de representación de estas construcciones se ha llevado a cabo un análisis crítico de las propuestas de representación de la GG (Bonet y Solà 1986; Haegeman 1991) y de la GD (Tesnière 1959; Mel'čuk 2003), que ha permitido determinar cómo estas estructuras deben ser tratadas en las gramáticas de Txala (Tabla 1).

	GG	GD	Txala
<i>Sustantivas</i>	$V_p \leftarrow \text{CONJ} \leftarrow V_s$	$V_p \leftarrow V_s \leftarrow \text{CONJ}$	$V_p \leftarrow \text{CONJ} \leftarrow V_s$
<i>Adverbiales</i>	$V_p \leftarrow \text{CONJ} \leftarrow V_s$	$V_p \leftarrow V_s \leftarrow \text{CONJ}$	$V_p \leftarrow \text{CONJ} \leftarrow V_s$
<i>Relativas</i>	$V_p \leftarrow \text{PR}_{\text{rel}} \leftarrow V_s$	$V_p \leftarrow V_s \leftarrow \text{PR}_{\text{rel}}$	$V_p \leftarrow V_s \leftarrow \text{PR}_{\text{rel}}$
<i>Pseudorelativas</i>	$V_p \leftarrow \text{PR}_{\text{rel}} \leftarrow V_s$	$V_p \leftarrow V_s \leftarrow \text{PR}_{\text{rel}}$	$V_p \leftarrow V_s \leftarrow \text{PR}_{\text{rel}}$
<i>S. con marcador elidido</i>	$V_p \leftarrow V_s$	$V_p \leftarrow V_s$	$V_p \leftarrow V_s$

Tabla 1. Resumen de la distribución de núcleos sintácticos de las subordinadas según GG, GD y Txala

3.3. La estructura de las comparaciones

Algunos estudios teóricos y descriptivos profundizan en el análisis de la estructura de las comparaciones, como son los trabajos de Solà (1972), Rivara (1990) y Gutiérrez Ordóñez (1994). No obstante, es escasa la caracterización de los fundamentos teóricos de esta estructura. Además, hay divergencias de opinión en la definición de las construcciones que realmente son comparativas y de los

elementos o marcadores que favorecen la lectura comparativa de una determinada construcción (Saragossà 2002: 3097).

Las construcciones comparativas tienen como finalidad comunicativa dar a conocer una información que es conocida por el receptor (Saragossà 2002: 3098). Esta definición es de carácter general y poco aporta en términos de la representación sintáctica de dicha construcción. De todas formas, puede ser desglosada y explicada más concretamente.

En la comparación, se expresa una información desconocida por el receptor contrastándola con información que ya posee. Ambas informaciones se manifiestan en la oración mediante dos términos comparados, X e Y, donde X es la incógnita que al ser comparada con la información conocida recogida en Y es resuelta (Gutiérrez Ordóñez 1994: 13). El término X aparece en primer lugar y se denomina antecedente, y el término Y se expresa a continuación del antecedente y, por ello, se le conoce como consiguiente. Los dos componentes de la comparación, X e Y, entran a funcionar dentro de la construcción comparativa gracias a que están introducidos por operadores gramaticales (p.e. *más, tan*, etc.) que los ponen en relación (18).

(18)

- i. Aquesta noia és tan alta com tu
‘Esta chica es tan alta como tú’
- ii. Aquesta noia és [antecedente tan alta] [consiguiente com tu]

A partir de la observación de (18), debemos cuestionarnos cómo debe ser la representación sintáctica en EsTxala y CaTxala. En realidad, existen dos análisis posibles (Fig. 18, Fig. 19):

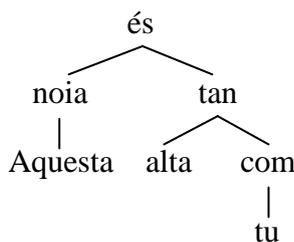


Fig. 18. Análisis de la construcción comparativa con el marcador como núcleo del ejemplo (18i) ‘Aquesta noia és tan alta com tu’ (‘Esta chica es tan alta como tú’)

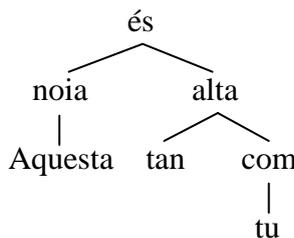


Fig. 19. Análisis de la construcción comparativa con el antecedente como núcleo

del ejemplo (18i) 'Aquesta noia és tan alta com tu' ('Esta chica es tan alta como tú')

En la Fig. 18, el análisis propuesto expresa implícitamente que el marcador funciona paralelamente como las conjunciones de las subordinadas sustantivas y adverbiales (§3.2.1) y, por lo tanto, es el introductor y el núcleo de la construcción comparativa. No obstante, al lado de ejemplos como (18i), existen otros ejemplos que plantean dificultades a esta propuesta de análisis (19).

(19) Tiene más dinero

En el ejemplo (19), se observa, por un lado, que el consiguiente no es expresado léxicamente y, por el otro, el introductor *más* actúa como un modificador del nombre *dinero*. De modo que, en realidad, el primer marcador de la comparación tiene el mismo valor que los cuantificadores (Brucart y Rigau 2002: 154) y, por lo tanto, modifica el núcleo del antecedente. En EsTxala y CaTxala, la estrategia para representar las construcciones comparativas se corresponde a este último análisis y, por lo tanto, los análisis propuestos son paralelos al análisis ilustrado en la Fig. 19, donde el antecedente es el núcleo de la construcción comparativa. A pesar de la validez de ambos análisis, desde la perspectiva del PLN, el tratamiento de las construcciones comparativas plantea graves problemas debido a la gran diversidad de estructuras y de elisiones (por ejemplo, elisión de toda la cláusula que expresa el consiguiente o parte de ella).

3.4. La conjunción coordinante como elemento nuclear

A lo largo de este artículo, se ha discutido sobre estructuras y construcciones sintácticas que tienen como rasgo común el hecho de que son relaciones estructurales jerárquicas entre un núcleo sintáctico y los nodos que dependen de él. Ahora bien, en cuanto a la coordinación, el fenómeno que se nos presenta es sustancialmente diferente, ya que generalmente en esta estructura se ha considerado que los constituyentes se encuentran alineados al mismo nivel, tal y como expone Trask (1992: 63):

coordinate structure *n.* A syntactic structure in which two or more constituents are joined ('conjoined') in such a way that each of them has an equal claim to be considered a head of that structure. [...]

No obstante, se ha discutido si la ordenación de los constituyentes se da al mismo nivel o, en realidad, se corresponde a una jerarquía (Bonet y Solà 1986: 321; López García 1999: 3513; Serra y Prunyonosa 2002: 2183).

Siguiendo a Bonet y Solà (1986: 349), desde una perspectiva generativista, la construcción coordinada consta de dos o más elementos que son equivalentes en cuanto a la función gramatical y están unidos al mismo nivel de la jerarquía estructural mediante un elemento de enlace (Dik 1968: 25). A partir de esa

definición, Bonet y Solà (1986) hipotetizan que la coordinación puede ser explicada mediante la regla (20), que favorece el análisis de la coordinación como una ordenación al mismo nivel de los constituyentes (Fig. 20).

$$(20) X \rightarrow X \text{ Coord } X^8$$

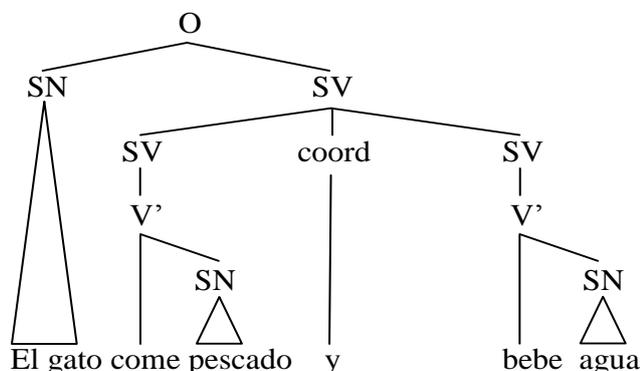


Fig. 20. Análisis de la construcción coordinada según la GG de la oración ‘El gato come pescado y bebe agua’

Al lado de la preposición y de la conjunción subordinante, la GD considera la conjunción coordinante una categoría intranuclear por su naturaleza conjuntiva y, por lo tanto, una clase de unidad funcional (Tesnière 1959: 53). Al no contener ninguna carga en cuanto a la función semántica, esta categoría no funciona como núcleo sintáctico de la coordinación (Mel’čuk 2003: 50), tal como se ilustra en Fig. 21.

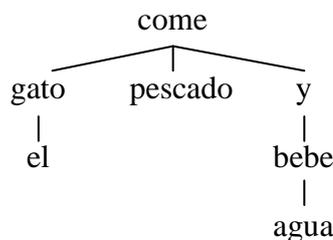


Fig. 21. Análisis de la construcción coordinada según la GD de la oración ‘El gato come pescado y bebe agua’

En este punto, EsTxala y CaTxala discrepan totalmente del análisis que propone la GD (Tesnière 1959; Mel’čuk 2003), ya que, como expone Bonet y Solà (1986: 349), la construcción coordinada no responde a una estructura jerárquica y la categoría que caracteriza esta construcción es la propia conjunción coordinada.

⁸ La regla expuesta a (20) prevé la posibilidad de piezas gramaticales léxicamente vacías, que pueden ser recuperadas mediante los elementos del primer coordinado (si los elementos vacíos se encuentran en el segundo coordinado), elementos externos a la construcción coordinada o elementos desplazados al margen derecho de la coordinación (en el caso que los elementos vacíos estén en el segundo coordinado).

Además, la estructura que propone la GD supone en numerosas ocasiones el nacimiento de ambigüedades estructurales (Mel'čuk 2003: 93), que a nuestro entender deben ser evitadas en el análisis automático del lenguaje. Por ejemplo, según los postulados de la GD, en la coordinación *wonderful photographs and paintings* ('fotografías y dibujos preciosos') hay implicadas dos interpretaciones. El adjetivo *wonderful* puede modificar sólo al nombre *photographs*, pero, a la vez, puede observarse que este adjetivo modifica a todo el grupo nominal coordinado, es decir, a *photographs y paintings*.

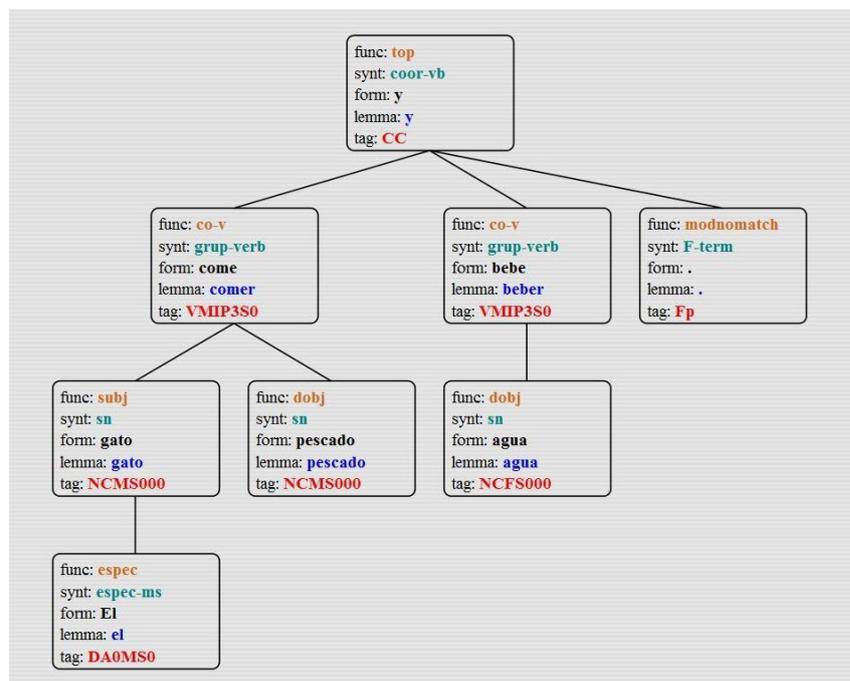


Fig. 22. Análisis de la construcción coordinada según EsTxala de la oración 'El gato come pescado y bebe agua'

De modo que el criterio establecido en ambas gramáticas para la coordinación no respalda las teorías de la GD. En otras palabras, el núcleo de toda coordinación debe ser la conjunción coordinada (Fig. 22).

4. CONCLUSIONES

En este artículo, hemos presentado los criterios adoptados para el tratamiento de diversas estructuras sintácticas en las gramáticas EsTxala y CaTxala. Hemos tratado aquellas construcciones o estructuras que desde una perspectiva teórica y computacional plantean diferentes soluciones de representación en función del formalismo lingüístico. Concretamente, hemos focalizado en la representación del sintagma preposicional y de las oraciones subordinadas (sustantivas, adverbiales, relativas, pseudorelativas y con elisión de marcador). A su vez, hemos abordado las estructuras comparativas y las coordinaciones.

Para llegar a un criterio claro en las gramáticas, hemos llevado a cabo una revisión crítica de las propuestas para este conjunto de construcciones y estructuras de los dos grandes formalismos lingüísticos implementados en análisis automático del lenguaje, la Gramática Generativa (Chomsky 1981) y la Gramática de Dependencias (Tesnière 1959; Mel'čuk 2003). Como consecuencia de la exposición de las diversas propuestas teóricas, se ha podido diseñar un repertorio de criterios lingüísticos con el fin de aportar soluciones fundamentadas sintácticamente y semánticamente.

A pesar de que la tarea de elaboración de criterios es abierta, consideramos que los criterios sintácticos que hemos establecido son una versión suficientemente elaborada, ya que tratan aquellos fenómenos sintácticos que han sido interpretados de manera distinta según la Gramática Generativa y la Gramática de Dependencias. No obstante, puesto que se trata de una tarea sin límites, en las siguientes etapas del proyecto se prevé la posibilidad de incorporar o modificar criterios según las necesidades de cada etapa. Estos criterios están implementados en las gramáticas que son de acceso abierto (<http://grial.uab.es/tools/download/>).

En paralelo, a este trabajo, hemos desarrollado un corpus de test para gramáticas automáticas que sigue los criterios propuestos y que incluye otras estructuras incluidas en EsTxala y CaTxala además de las presentadas en este artículo. Dicho repertorio es abierto y consultable en línea (http://161.116.36.206/~publicacions/articlesDEPGRAM/repertori_avaluacio_text.pdf).

La definición del repertorio de criterios sintácticos permite plantear como línea de investigación futura la evaluación de la calidad del recurso desarrollado, las gramáticas EsTxala y CaTxala, y, en último término, la evaluación de los criterios sintácticos propuestos. Para llevar a cabo esta tarea, estamos diseñando un protocolo de evaluación para dichas gramáticas que ponga a prueba la calidad del conocimiento lingüístico contenido en las gramáticas, ya sea información sintáctica (anidación del sintagma preposicional, formación de oraciones complejas, subcategorización verbal, estructuras coordinadas, etc.) como información semántica (preferencias selectivas, rasgos de *Top Concept Ontology*).

BIBLIOGRAFÍA

- ALARCOS, E. (1994), *Gramática de la lengua española*, Madrid, Espasa Calpe, Real Academia Española.
- ALSINA, À., BADIA, T., BOLEDA, G., BOTT, S., GIL, À., QUIXAL, M. y VALENTÍN, O. (2002), "CATCG: Un sistema de análisis morfosintáctico para el catalán", en *Procesamiento del Lenguaje Natural*, 29, 309-310.
- ATSERIAS, J., COMELLES, E. y MAYOR, A. (2005), "Txala un analizador libre de dependencias para el castellano", en *Procesamiento del Lenguaje Natural*, 35:455-456.

- BICK, E. (2006). "A constraint grammar-based parser for Spanish", en *Proceedings of TIL 2006, 4th Workshop on Information and Human Language Technology*.
- BADIA I MARGARIT, A. M. (1962), *Gramàtica catalana*, Madrid, Gredos.
- BONET, S. y SOLÀ, J. (1986), *Sintaxi generativa catalana*, Barcelona, Biblioteca Universitària, Enciclopèdia Catalana.
- BONET, S. (2002), "Les subordinades substantives", en *Gramàtica del Català Contemporani*, Solà, J., Lloret, M.R., Mascaró, J. y Pérez Saldanya, M. (dirs.), vol. 3, Barcelona, Empúries, 2321-2387.
- BOSQUE, I. y DEMONTE, V. (dirs.) (1999), *Gramàtica descriptiva de la lengua española*, Madrid, Espasa.
- BRUCART, J. M. (1999), "La estructura del sintagma nominal: las oraciones de relativo", en *Gramàtica descriptiva de la lengua española*, Bosque, I. y Demonte, V. (dirs.), vol. 1., Madrid, Espasa, 395-522.
- BRUCART, J. M. y RIGAU, G. (2002), "La quantificació", en *Gramàtica del Català Contemporani*, Solà, J., Lloret, M.R., Mascaró, J. y Pérez Saldanya, M. (dirs.), vol. 2, Barcelona, Empúries, 1517-1589.
- CALVO, H. y GELBUKH, A. (2006), "DILUCT: An open-source Spanish dependency parsers based on rules, heuristics, and selectional preferences", en *Natural Language Processing and Information Systems. Lecture Notes in Computer Science*, Kop, C., Fliedl, G., Mayr, H.C. y Métais, E. (eds.), Berlin, Springer-Verlag, vol. 3999, 164-175.
- CASTELLÓN, I., CIVIT, M. y ATSERIAS, J. (1998), "Syntactic parsing of unrestricted spanish text", en *First International Conference on Language Resources and Evaluation*, España, Granada.
- CHOMSKY, N. (1957), *Syntactic Structures*, The Hague, Mouton de Gruyter.
- _____ (1981), *Lectures on Government and Binding: The Pisa Lectures*, The Hague, Mouton de Gruyter.
- COLLINS, M. (2000), "Discriminative reranking for natural language parsing", en *Proceedings of the Seventeenth International Conference on Machine Learning (ICML 2000)*, 175-182.
- COMELLES, E., ARRANZ, V. y CASTELLÓN, I. (2010), "Constituency and Dependency Parsers Evaluation", en *Procesamiento del Lenguaje Natural*, 45: 59-66.
- DE MARNEFFE, M.C., MACCARTNEY, B. y MANNING, C. D. (2006), "Generating typed dependency parses from phrase structure parses", en *Proceedings of the Fifth Conference on International Language Resources and Evaluation (LREC'06)*, 449-454.
- DELBECQUE, N. y LAMIROY, B. (1999), "La subordinación sustantiva: las subordinadas enunciativas en los complementos verbales", en *Gramàtica descriptiva de la lengua española*, Bosque, I. y Demonte, V. (dirs.), vol. 2., Madrid, Espasa, 1965-2081.
- DIK, S. C. (1968), *Coordination. Its implications for the theory of general linguistics*, Amsterdam, North-Holland.
- FABRA, P. (1918³), *Gramàtica catalana*, Barcelona, Institut d'Estudis Catalans.
- FERRÁNDEZ, A., PALOMAR, M. y MORENO, L. (2000), "Slot unification grammar and anaphora resolution", en *Recent Advances in Natural Language Processing II. Selected papers from RANLP 1997*, Nicolov, N. y Mitkov, R. (eds.), John Benjamins, 155-166.
- GAMALLO, P. y GONZÁLEZ, I. (2009), "Una gramàtica de dependencias basada en patrones de etiquetas", en *Procesamiento del Lenguaje Natural*, 43, 315-323.
- GUTIÉRREZ ORDÓÑEZ, S. (1994), *Estructuras comparativas*, Madrid, Arco/Libros.

- HAEGEMAN, L. (1991), *Introduction to Government and Binding Theory*, Oxford, Blackwell.
- HERNANZ, M. LI. (2002), "L'oració", en *Gramàtica del Català Contemporani*, Solà, J., Lloret, M.R., Mascaró, J. y Pérez Saldanya, M. (dirs.), vol. 2, Barcelona, Empúries, 993-1073.
- HERNANZ, M. LI. y BRUCART, J. M. (1987), *La sintaxis. Principios teóricos. La oración simple*, Barcelona, Editorial Crítica.
- KOO, T., CARRERAS, X. y COLLINS, M. (2008), "Simple semi-supervised dependency parsing", en *Proceedings of 46th Annual Meeting on Association for Computational Linguistics (ACL'08)*, 595-603.
- LLOBERES, M., CASTELLÓN, I. y PADRÓ, LI. (2010), "Spanish FreeLing Dependency Grammar", en *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Calzolari, N. et al. (eds.), Valletta, Malta.
- LIN, D. (1998), "Dependency-based evaluation of MINIPAR", en *Proceedings of the LREC Workshop on the Evaluation of Parsing Systems*, España, Granada.
- LÓPEZ GARCÍA, Á. (1999), "Relaciones paratácticas e hipotácticas", en *Gramática descriptiva de la lengua española*, Bosque, I. y Demonte, V. (dirs.), vol. 2., Madrid, Espasa, 3507-3547.
- MAGERMAN, D. M. (1995), "Statistical decision-tree models for parsing", en *Proceedings of the 33rd Annual Meeting on Association for Computational Linguistics (ACL'95)*, 276-283.
- MARIMÓN, M. (2010), "The Spanish Resource Grammar", en *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)*, Calzolari, N. et al. (eds.), Valletta, Malta.
- MEL'ČUK, I. A. (1988), *Dependency Syntax: Theory and Practice*, Albany-New York, The SUNY Press.
- _____ (2003), "Dependency in Natural Language", en *Dependency in Linguistic Description. Studies in Language Companion Series*, Polguère, A. y Mel'čuk (eds.), I.A., Amsterdam-Philadelphia, John Benjamins Publishing Co., 1-110.
- MORENO CABRERA, J.C. (1991²), *Curso universitario de lingüística general: Teoría de la gramática y sintaxis general*, Madrid, Síntesis.
- NIVRE, J. (2006). *Inductive Dependency Parsing*, en *Text, Speech and Language Technology*, vol. 34, Dordrecht, Springer-Verlag, 2006.
- PADRÓ, LI., REESE, S., LLOBERES, M. y CASTELLÓN, I. (2010), "Freeling 2.1: Five years of open-source language processing tools", en *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Calzolari, N. et al. (eds.), Valletta, Malta.
- RIVARA, R. (1990), *Le système de la comparaison. Sur la construction du sens dans les langues naturelles*, Paris, Les Éditions de Minuit.
- SANCHO CREMADES, P. (2002), "La preposició i el sintagma preposicional", en *Gramàtica del Català Contemporani*, Solà, J., Lloret, M.R., Mascaró, J. y Pérez Saldanya, M. (dirs.), vol. 2, Barcelona, Empúries, 1689-1796.
- SARAGOSSÀ, A. (2002), "Les construccions comparatives i les oracions consecutives", en *Gramàtica del Català Contemporani*, Solà, J., Lloret, M.R., Mascaró, J. y Pérez Saldanya, M. (dirs.), vol. 3, Barcelona, Empúries, 3095-3171.
- SCHMID, H. (1994), *TreeTagger - a language independent part-of-speech tagger*, <http://www.ims.uni-stuttgart.de/Tools/DecisionTreeTagger.html>.

- SERRA, E. y PRUNYONOSA, T. (2002), "La coordinació", en *Gramàtica del Català Contemporani*, Solà, J., Lloret, M.R., Mascaró, J. y Pérez Saldanya, M. (dirs.), vol. 3, Barcelona, Empúries, 2181-2245.
- SLEATOR, D. D.K. y TEMPERLEY, D. (1991), "Parsing english with a link grammar", en *Third International Workshop on Parsing Technologies*, Tilburg, The Netherlands y Durbuy, Belgium.
- SOLÀ, J. (1972²), *Estudis de sintaxi catalana*, Barcelona, Edicions 62.
- _____ (2002), "Les subordinades de relatiu", en *Gramàtica del Català Contemporani*, Solà, J., Lloret, M.R., Mascaró, J. y Pérez Saldanya, M. (dirs.), vol. 3, Barcelona, Empúries, 2455-2565.
- SOLÀ, Joan, LLORET, M. Rosa, MASCARÓ, Joan y PÉREZ SALDANYA, Manuel (dirs.) (2002), *Gramàtica del Català Contemporani*, Barcelona, Empúries.
- TAPANAINEN, P. y JÄRVINEN, T. (1998), "Towards an implementable dependency grammar", en *Proceedings of the ACL Workshop on Processing of Dependency-Based Grammars*, 1-10.
- TESNIÈRE, L. (1959), *Elements de syntaxe structurale*, Paris, Klincksieck.
- TRASK, R. L. (1992), *A dictionary of grammatical terms in Linguistics*, London-New York, Routledge.
- VILLALBA, X. (2002), "La subordinació", en *Gramàtica del Català Contemporani*, Solà, J., Lloret, M.R., Mascaró, J. y Pérez Saldanya, M. (dirs.), vol. 3, Barcelona, Empúries, 2247-2319.
- YAMADA, H. y MATSUMOTO, Y. (2003), "Statistical dependency analysis with support vector machines", en *Proceedings of 8th International Workshop on Parsing Technologies*, 195-206.